Downloaded from orbit.dtu.dk on: May 07, 2023

DTU Library

Discovery and description of complete ammonium oxidizers in groundwater-fed rapid sand filters

Palomo, Alejandro

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Palomo, A. (2017). *Discovery and description of complete ammonium oxidizers in groundwater-fed rapid sand filters*. Department of Environmental Engineering, Technical University of Denmark (DTU).

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Discovery and description of complete ammonium oxidizers in groundwater-fed rapid sand filters

Alejandro Palomo-González

PhD Thesis
June 2017

DTU Environment
Department of Environmental Engineering
Technical University of Denmark

Alejandro Palomo-González

Discovery and description of complete ammonium oxidizers in groundwater-fed rapid sand filters

PhD Thesis, June 2017

The synopsis part of this thesis is available as a pdf-file for download from the DTU research database ORBIT: http://www.orbit.dtu.dk.

Address: DTU Environment
Department of Environmental Engineering
Technical University of Denmark
Bygningstorvet, Bygning 115
2800 Kgs. Lyngby
Denmark
Phone reception: +45 4525 1600
Fax: +45 4593 2850
Homepage: http://www.env.dtu.dk
E-mail: reception@env.dtu.dk
Cover: GraphicCo

Preface

This thesis is based on the work carried out at the Technical University of Denmark, Department of Environmental Engineering and the Department of Bio and Health Informatics from January 2014 to May 2017. This thesis was part of the Mermaid, ITN-EU-FP7 funded by the People Programme (Marie Skłodowska-Curie Actions). The research was performed under the main supervision of Professor Thomas Sicheritz-Pontén (DTU Bioinformatics) and Professor Barth F. Smets (DTU Environment), and co-supervision of Simon Rasmussen (DTU Bioinformatics).

I **Palomo, A.**, Fowler, S.J., Rasmussen, S., Sicheritz-Pontén, T., Smets, B.F. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *ISME J.* **10,** 2569–2581 (2016).

II **Palomo, A.**, Pedersen A.G., Fowler, S.J., Dechesne, A., Sicheritz-Pontén, T., Smets, B.F. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox *Nitrospira*. *Submitted manuscript*.

III Fowler, S.J., **Palomo, A.**, Smets, B.F. Comammox *Nitrospira* are the dominant ammonia oxidizers in diverse rapid sand filter communities. *Submitted manuscript*.

IV **Palomo, A.**, Fowler, S.J., Nemer, I.M., Smets, B.F. Examining differential abundance in rapid sand filter microbial communities after short-term ammonium loading-disturbances. *Manuscript in preparation*.

TEXT FOR WWW-VERSION (without papers)

In this online version of the thesis, paper **I-IV** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from DTU Environment, Technical University of Denmark, Bygningstorvet, Bygning 115, 2800 Kgs. Lyngby, Denmark, info@env.dtu.dk

In addition, the following co-authored publications, not included in this thesis, were also concluded during this PhD study:

- Dechesne, A., Musovic, S., **Palomo, A.**, Diwan, V., Smets, B.F. (2016). Underestimation of ammonia-oxidizing bacteria abundance by amplification bias in amoA-targeted qPCR. *Microb Biotechnol* **9**: 519–524.
- Kinnunen, M., **Palomo, A.**, Fowler, S.J., Albrechtsen, H-J., Dechesne, A., Smets, B.F. Changes in well-stablished rapid sand filter community as a response to long-term exclusive nitrite loading. *Manuscript in preparation*.
- Azevedo, D., **Palomo, A.**, Smets, B.F. Metagenomics analysis of a single-stage nitritation/anammox sequencing batch reactor. *Manuscript in preparation*.

This PhD study also contributed to international conferences with the following proceeding and conference papers:

- **Palomo, A.**, Fowler, S.J., Rasmussen, S., Sicheritz-Pontén, T., Smets, B.F. Metagenomics and single-cell genomics reveal high abundance of comammox *Nitrospira* in a rapid gravity sand filter treating groundwater. 16th International Symposium on Microbial Ecology, 2016, Montreal, Canada
  Poster Presentation
- **Palomo, A.**, Fowler, S.J., Rasmussen, S., Schramm, A., Sicheritz-Pontén, T., Smets, B.F. Investigating comammox *Nitrospira* in rapid sand filters via metagenomics and single-cell genomics. Microbial Ecology in Water Engineering & Biofilms 2016, Copenhagen, Denmark.
  Oral Presentation
- **Palomo A**, Gülay A, Rasmussen S, Sicheritz-Pontén T, Smets B.F. Taxonomic and Metagenomic profiling of rapid sand filter microbiome reveals a high *Nitrospira* incidence. 4th International Conference on Nitrification and Related Processes (ICON), 2015, Edmonton, Canada
  Oral presentation
- **Palomo A**, Rasmussen S, Sicheritz-Pontén T, Smets B.F. Metagenomic analysis of microbial communities in rapid sand filter treating groundwater. Community diversity and metabolic potential. 6th Congress of European Microbiologists FEMS 2015, Maastricht, The Netherlands
  Poster presentation

- Gülay A, **Palomo A**, Musovic S, Albrechtsen, H-J., Smets B.F. Diversity and metabolic potential of the microbial communities in rapid sand filters at Danish waterworks. Danish Water Forum 9th 2014, Copenhagen, Denmark.
  Oral presentation by Arda Gülay
- Musovic S, **Palomo A**, Diwan V, Dechesne A, Smets B.F. qPCR quantification of ammonia oxidizing bacteria: What should the target be? (2014). Danish Microbiological Society Annual Congress 2014, Copenhagen, Denmark
  Poster presentation
- Smets B.F., Gülay A, **Palomo A,** Fowler, S.J., Sicheritz-Pontén T, Diversity, structure, and novel physiologies in microbial communities in rapid sand filters. 7th Congress of European Microbiologists FEMS 2017, Valencia, Spain
  Accepted for Oral presentation by Barth F. Smets
- **Palomo, A.**, Fowler, S.J., Pedersen, A.G., Dechesne, A., Sicheritz-Pontén, T., Smets, B.F. Niche differentiation and evolution of comammox *Nitrospira* through a comparative genomics analysis. 5th International Conference on Nitrification and Related Processes (ICON), 2017, Vienna, Austria
  Accepted for Oral presentation

Acknowledgements

First, I would like to thank my two main supervisors, Barth F. Smets and Thomas Sicheritz-Pontén for their professional guidelines, valuable feedback, and for giving me the opportunity to do this PhD. I will also like to show my gratitude to my co-supervisor Simon Rasmussen and to my former co-supervisor Jacob Bælum for providing me their expertise whenever I have required it. Thanks to my co-authors Jane Fowler, Arnaud Dechesne, Arda Gülay and Anders Gorm Pedersen for the great collaboration and fruitful discussions.

I would like to thank the Mermaid project and the Marie Skłodowska-Curie action for the high quality of the training and the great organization. I have been able to gain a lot of expertise from very different disciplines which will help me in my future career. It has been an enormous privilege to meet with fantastic colleagues and friends every six months all around Europe.

I wish to thank the former and present members of the Metlab research group. It has been a pleasure to work and share this time with all of you. Thanks to the members of the Metagenomics group who helped me to learn some of the required skills to carry out my PhD. I am grateful for the assistance of the laboratory technicians, the administration staff and the IT group from DTU environment and DTU Bioinformatics. I would like to thank Marlene Mark Jensen for the Danish version of the summary, and Philip Binning for the photo of the sand filter used in the cover of this thesis. Thanks to the waterworks involved in this study for allowing me to sample the filters and to the workers for their patience and assistance. A special thanks to my students Queralt, Majken, Helene, Ibrahim, Nikola and Daniela for their enthusiasm and their hard work.

Thanks to all my colleague and friends who have made me enjoy these three and a half years at DTU: Alberto, Bas, Borja, Carson, Dongah, Dorka, Elena, Fabio, Florian, Fynn, Jan, Katerina, Marta, Pauly, Pedram, Sara, Tonio, Uli, Vaibhav, and many others; and especially to Carlos and Ma, for all the moments lived together and sharing this journey with me, it will have not been the same without you.

Finally, I would really like to thank my family, who has been always there supporting me, and trying to do their best for my happiness and my future.

Summary

Microbial communities are directly linked with process performance in several engineered systems. In the last century, intense study of microorganisms has contributed to optimize important environmental biotechnologies such as the activated sludge process or anaerobic digestion. However, less attention has been paid to the role of microorganisms in drinking water treatment technologies. In contrast, much effort has been devoted to eliminate potential pathogens in the drinking water treatment and supply systems. Nevertheless, the role of microbes in some drinking water treatments systems as biological filtration has long been acknowledged and recently been investigated. Biological filtration technology is widely used around the world and is especially important in Denmark as groundwater is the main source water for drinking water production. Because the groundwater has a relative high-quality, aeration followed by biological filtration is the only required treatment before distribution. In the last years, the microbial communities in rapid gravity sand filters, the typical biological filter used in Denmark, have been characterized, but little knowledge had been required about their physiological activity and roles in compound removal from the source water.

This PhD project focused on a comprehensive investigation of the microbial communities in rapid sand filters beyond their purely taxonomical identification. For this purpose, samples collected from a rapid sand filter were subjected to metagenomics analysis and genome recovery to identify the genetic capacities of the dominant types in the microbial community. Fourteen near-complete population genomes representing the dominant community were recovered comprising the capacity to grow on the typical compounds found in groundwater. The identified population genomes contained capabilities to oxidize ammonium, nitrite, methane, hydrogen sulfide, iron and manganese as well as to assimilate organic compounds. A composite population genome was assigned to *Nitrospira*. This genus had previously been found in multiple rapid sand filters at an unexplained high abundance. *Nitrospira* spp. are known to perform the second step of nitrification: oxidation of nitrite to nitrate. The two-step nitrification process disclosed at the end of the 19th century was assumed to be carried out by two different functional groups, ammonia oxidizing prokaryotes and nitrite oxidizing bacteria. Strikingly, the *Nitrospira* composite population genome not only contained the genes to oxidize nitrite to nitrate, but also the genetic potential to execute the first step of ni-

trification. Exhaustive bioinformatics investigation ruled out the possibility of genomic contamination and confirmed that the *Nitrospira* composite population genome harboured the complete ammonium oxidation (comammox) pathway. At the same time, evidence of a single microbe's capacity to carry out complete nitrification was obtained by three other groups; in all cases the comammox type belonged to the *Nitrospira* genus.

To further investigate the genomic capacities of comammox *Nitrospira*, the *Nitrospira* composite genome was separated into individual population genomes using a differential coverage binning approach. As a result, five individual genomes were recovered, four of them containing the complete ammonium oxidation pathway. These genomes together with 11 high-quality publically available *Nitrospira* genomes (seven comammox and four strict nitrite oxidizers) were subject to a comparative genomics analysis. This examination showed specific genomic features for comammox, strict nitrite oxidizers and the two comammox clades. Thus, comammox *Nitrospira* harbour a higher variety of genes related to adaptation to nutrient-limited environments. The two comammox clades differ in their ammonium uptake affinity systems. Additionally, comammox *Nitrospira* genomes lack the genetic capacity to use nitrite as the only nitrogen source.

The evolutionary history of comammox *Nitrospira* was also examined based on protein dissimilarity, gene arrangement and reconciliation analysis. We detected a high probability of horizontal gene transfer events from betaproteobacterial ammonia oxidizers to comammox *Nitrospira* for genes belonging to the ammonium oxidation pathway as well as from comammox clade B to clade A for a subset of genes.

I investigated the abundance of comammox *Nitrospira* in rapid sand filters at 12 different waterworks in Denmark. As these new microorganisms are taxonomically similar to strict *Nitrospira* nitrite oxidizers, we developed specific primers to exclusively target comammox based on their gene encoding the ammonia monooxygenase subunit A. With these primers, we detected comammox *Nitrospira* as the dominant nitrifier in the biofilters with an abundance typically one order of magnitude higher than canonical ammonium oxidizing prokaryotes.

Lastly, I carried out lab-scale experiments with filter material from the top and bottom layers of a rapid sand filter containing different proportions of

comammox *Nitrospira*, and strict nitrite and ammonia oxidizing prokaryotes under different loading conditions. Specifically, I exposed the filter material to distinct ammonium loading, under presence or absence of external carbon source as well as under oxygen limitation. In relation to the nitrifying community three main findings were made: (i) simultaneous growth of comammox *Nitrospira* and ammonium oxidizing prokaryotes; (ii) lower fitness of ammonium oxidizing archaea at higher temperatures; (iii) selection of comammox clade A over clade B at increasing ammonium loadings at reference temperature.

Overall, this PhD has provided insights into the genomic capabilities of the main types in the microbial community of a groundwater-fed biological filter. Moreover, the previously observed high abundances of *Nitrospira* spp. in rapid sand filters, has now been explained, by the discovery of complete ammonium oxidizing (comammox) *Nitrospira* from metagenomics analysis. In addition, this thesis presents the first extensive analysis of the genomic capabilities of comammox *Nitrospira* compared to canonical ammonium and nitrite oxidizers.

Dansk sammenfatning

Det mikrobielle samfund er direkte koblet til procesydelse i flere teknologiske systemer. Igennem det sidste århundrede har intensive undersøgelser af mikroorganismer bidraget til optimering af vigtige miljøteknologier, såsom aktiveret slam eller anaerob forrådnelse. Man kender dog kun meget lidt til mikroorganismernes rolle i vandforsyningsanlæg. Der har været en stor indsats for at fjerne potentielle patogener i vandforsyningsanlæg og vandforsyningssystemerne. Ikke desto mindre har mikroorganismernes rolle i nogle vandforsyningsanlæg, såsom biologiske vandrensningsfiltre, længe været anerkendt og er for nylig blevet undersøgt. Biologisk filter teknologi er udbredt i hele verden og er særligt brugt i Danmark, da næsten al dansk drikkevand kommer fra grundvand. Idet grundvandet har en relativ høj drikkevandskvalitet, er beluftning efterfulgt af biologiske sandfiltre den eneste behandlingsform inden vandet ledes ud i forsyningssystemerne. I de seneste år er det mikrobielle samfund i hurtig sandfiltre, der typisk anvendes i Danmark, bl evet karakteriseret, men der er kun begrænset viden om deres fysiologiske aktivitet og roller i fjernelsen af forskellige kemiske forbindelser fra grundvand.

Denne ph.d.-afhandling omhandler en omfattende undersøgelse af de mikrobielle samfund i hurtig sandfiltre ud over deres taxonomiske identifikation. Til dette formål blev prøver indsamlet fra et hurtig sandfilter. Ved hjælp af metagenomiske analyser og gendannelse af genomer blev de genetiske kapaciteter af de mest dominerende typer i det mikrobielle samfund identificeret. Fjorten næsten komplette populations-genomer, der repræsenterer det dominerende mikrobielle samfund, blev kortlagt. Disse genomer indikerede mikroorganismernes evne til at vokse på typiske kemiske forbindelser, der forekommer i grundvand. De identificerede populations-genomer indeholdt genkomponenter, der koder for evnen til at oxidere ammonium, nitrit, metan, sulfid, reduceret jern og mangan samt evnen til at assimilere organiske forbindelser. Et sammensat populationsgenom kunne tildeles slægten *Nitrospira*. Denne slægt er tidligere blevet fundet i flere hurtig sandfiltre med en uforklarlig høj relativ tilstedeværelse. *Nitrospira* spp. er kendt for at være involveret i det andet trin af nitrifikationsprocessen: oxidation af nitrit til nitrat. Nitrifikation er en to-trins proces - beskrevet i slutningen af 1800-tallet - hvor processen blev antaget at være udført af to forskellige funktionelle grupper, de ammoniak-oxiderende prokaryoter og nitrit-oxiderende bakterier. Det er dog påfaldende, at *Nitrospira* populationsgenomet ikke kun indeholdt gener-

ne til at oxidere nitrit til nitrat, men også det genetiske potentiale til at udføre det første trin af nitrifikation. En omfattende bioinformatisk analyse udelukkede muligheden for en genom forurening og bekræftede at *Nitrospira*-tilknyttede populationensgenom indeholdt hele ammonium-oxidationsprocessen til nitrite (comammox). Samtidig blev det påvist at en enkelt mikroorganismes evne til at udføre komplet nitrifikation blev opnået af tre andre grupper, som alle tilhørte comammox typen af *Nitrospira* slægten.

For yderligere at kunne undersøge gen kapaciteterne af comammox *Nitrospira* blev det *Nitrospira*-sammensatte genom separeret i individuelle populationsgrupper ved anvendelse af en metode, der på engelsk kaldes "differential coverage binning approach". Ved hjælp af denne metode blev fem individuelle genomer genskabt, hvor fire indeholdt den komplette ammonia-oxidationsvej. Disse genomer blev sammen med 11 offentligt tilgængelige *Nitrospira*-genomer af høj kvalitet (syv comammox og fire kun nitritoxiderende) underkastet en genomisk analyse der tilader sammenligning imellem disse. Denne undersøgelse viste specifikke gen-træk for comammox, nitrit-oxiderende bakterier og de to comammox-relaterede-clader. Comammox Nitrospira besad flere forskellige gener, der relaterer til en tilpasning til næringsfattige miljøer. De to comammox klader adskiller sig i deres ammoniumoptagelsesaffinitetssystemer. Derudover mangler comammox *Nitrospira*-genomerne den genetiske kapacitet til at anvende nitrit som den eneste kvælstofkilde.

Baseret på proteinforskellighed, genarrangement og afstemningsanalyser blev den evolutionære historie af comammox *Nitrospira* undersøgt. Vi opsporede en høj sandsynlighed for horisontal genoverførselshændelser fra ammoniak-oxiderende bakterier, der hører til betaproteobakterier, til comammox *Nitrospira* af gener tilhørende ammoniumoxidationsvejen såvel som fra comammox-klade B til clade A af en delmængde af gener.

Jeg har desuden undersøgt antallet af comammox *Nitrospira* i hurtig sandfiltre på 12 forskellige vandværker i Danmark. Da disse nye mikroorganismer er taxonomisk identiske med nitritoxiderende *Nitrospira*, har vi udviklet specifikke primer, der udelukkende binder til comammox-genet, der koder for ammoniakmonooxygenase subunit A. Med disse primer opdagede vi at comammox *Nitrospira* var den dominerende nitrificerende bakterie i sandfilterne med en tilstedeværelse, der typisk var en størrelsesorden højere end de typiske ammonia-oxidiserende prokaryoter.

Jeg har desuden udført laboratorieeksperimenter med sand fra de øverste og nederste lag af hurtig sandfilter, der indeholdt forskellige antal af comammox *Nitrospira*, nitrit- og ammoniak-oxiderende prokaryoter udsat for forskellige belastningsbetingelser. Jeg eksponerede sandfiltermaterialet for forskellige ammonium belastninger, med eller uden ekstern kulstofkilde samt under iltbegrænsning. I forhold til det nitrificerende samfund er der tre primære resultater: (i) vækst af både comammox *Nitrospira* og ammonium-oxiderende prokaryoter; (ii) dårligere vækst af ammonium-oxiderende archaea ved højere temperaturer; (iii) selektion af comammox-klade A over klade B ved stigende ammoniumbelastninger ved reference temperatur.

Samlet set har dette ph.d. studie givet indsigt i de genomiske potentialer hos de vigtigste mikroorganismer i det mikrobielt samfund fundet i et biologisk sandfilter, der behandler grundvand. Desuden er de tidligere undersøgelser, der observerede et uforklarligt højt antal af *Nitrospira* spp. i hurtig sandfiltre, nu blevet belyst ved opdagelsen af komplet ammonia-oxidiserende (comammox) *Nitrospira* ud fra metagenomiske - analyser. Tilmed præsenterer denne afhandling den første omfattende analyse af de genomiske evner af comammox *Nitrospira* i forhold til de allerede kendte ammonia- og nitritoxiderende microorganismer.

Table of contents

Abbreviations

AO	Ammonia oxidizer
AOA	Ammonia oxidizing archaea
AOB	Ammonia oxidizing bacteria
AMO	Ammonia monooxygenase
anammox	Anaerobic oxidation of ammonium
AOP	Ammonia oxidizing prokaryotes
CDS	Coding sequence
comammox	Complete ammonia oxidizer
DNA	Deoxyribonucleic acid
DNRA	Dissimilatory nitrite reduction to ammonium
FISH	Fluorescence in situ hybridization
GAC	Granular activated carbon
HAO	Hydroxylamine dehydrogenase
NOB	Nitrite oxidizing bacteria
NXR	Nitrite oxidoreductase
PC	Protein cluster
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
SSF	Slow sand filters
TCA	Tricarboxylic acid
RNA	Ribonucleic acid
RSF	Rapid sand filters
rTCA	Reductive (reverse) tricarboxylic acid
WWTP	Wastewater treatment plant

1 Introduction

Nitrogen (N) is an essential element required for life as it is part of proteins and nucleic acids. N exists in multiple oxidation states, ranging from -3 (ammonium/ammonia) to +5 (nitrate), as well as several chemical forms. The different transformations and transitions between different nitrogen species form the nitrogen cycle (Figure 1.1). Most of these reactions are driven by microorganisms. Traditionally, the biogeochemical nitrogen cycle was composed of the following processes: N_2 fixation, nitrification, denitrification, dissimilatory nitrite reduction to ammonium (DNRA), and Org-NH_2 mineralization and assimilation of NH_4^+ (Figure 1.1). Nitrogen fixation plays a central process in the cycle because, although N is very abundant in the atmosphere (approximately 78%), most of the organisms require fixed forms of N.

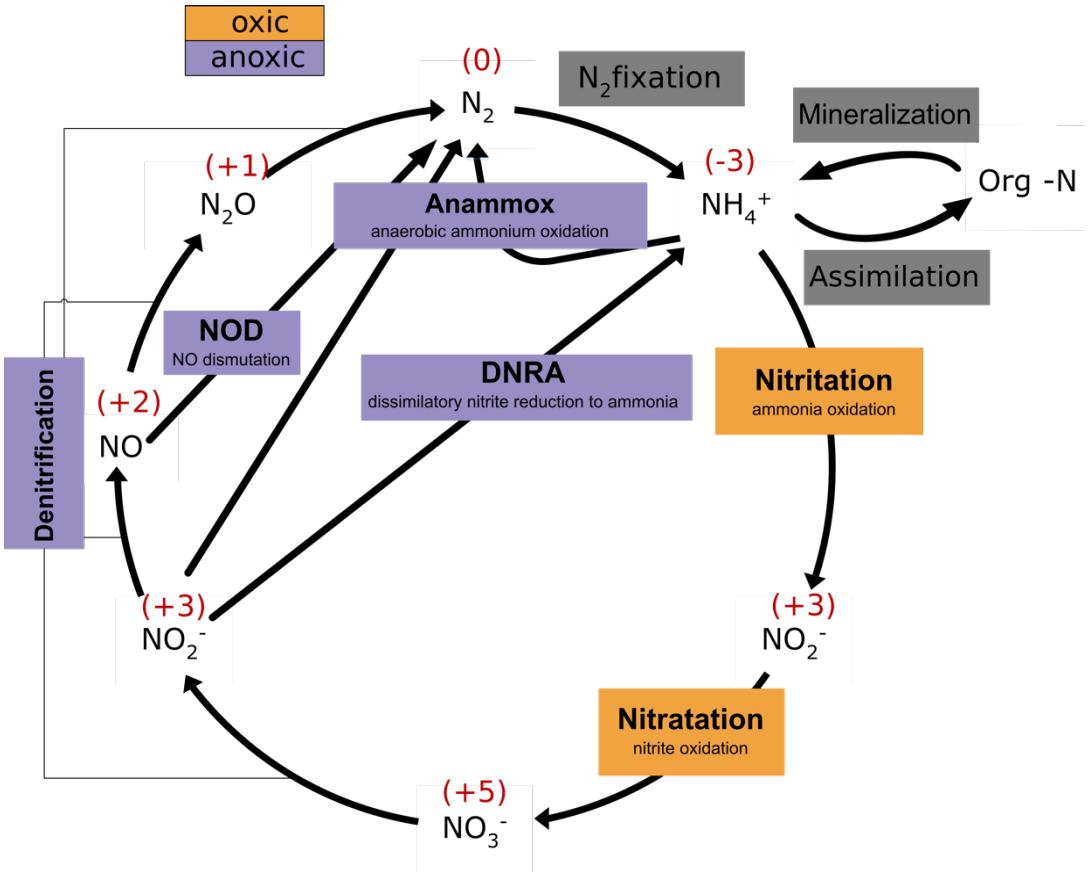

Figure 1.1. Major processes of the nitrogen cycle including the oxidation state of each intermediate.

Free-living or symbiotic bacteria and archaea carry out this process, producing ammonia. The result ammonia is either assimilated by organisms or converted to nitrate via nitrite in the presence of oxygen. This last transfor-

mation, known as nitrification, is a two-step process driven by ammonia oxidizing prokaryotes (AOP) and nitrite oxidizing bacteria (NOB). Nitrate is reduced to dinitrogen gas through denitrification or to ammonium via the DNRA, closing the cycle. Both denitrification and DNRA occurr in absence of oxygen.

1.1 New discoveries in the nitrogen cycle

This classical picture of the nitrogen cycle has drastically changed in the last decades. In the early 1950s, hydroxylamine (NH_2OH) was discovered as an intermediate in the first step of nitrification in *Nitrosomonas* spp. (Lees, 1952; Hofman and Lees, 1953) (Figure 1.2). In 1977, Broda proposed that two microbes related to the nitrogen cycle were missing (Broda, 1977). One of the organisms would oxidize ammonium to gaseous nitrogen with nitrate or nitrite as electron acceptor. The another organism would use inorganic nitrogen compounds as the electron donor for light-driven CO_2 fixation in anoxygenic phototrophs. Almost two decades later, his first hypothetical reaction was discovered, the anaerobic oxidation of ammonium (anammox) (Mulder *et al.*, 1995; Van de Graaf *et al.*, 1995) (Figure 1). Specific bacteria belonging to the order *Candidatus* Brocadiales grow slowly (doubling time ca. 7–22 days) from the transformation of ammonium and nitrite into dinitrogen gas in the absence of oxygen, and with CO_2 as the only carbon source. These microorganisms use a specific pathway with nitric oxide and hydrazine as intermediates (Kartal *et al.*, 2011). Anammox process is responsible for approximately 50% of nitrogen loss in marine and freshwater environments (Kuypers *et al.*, 2005; Schubert *et al.*, 2006). Ten more years had to wait for Broda's second prediction, when a *Thiocapsa* isolate was observed to carry out anoxygenic photosynthesis with the use of nitrite (instead of the ammonium, as anticipated by Broda) as electron donor combined with autotrophic CO_2 fixation (Griffin *et al.*, 2007). So far this metabolism seems to be restricted to few strains belonging to the genera *Rhodopseudomonas* (Alphaproteobacteria) and *Thiocapsa* (Gammaproteobacteria) (Schott *et al.*, 2010). Another N-related process recently disclosed is the anaerobic oxidation of ammonium coupled with iron reduction (feammox) (Clément *et al.*, 2005; Sawayama, 2006). Studies have linked this reaction with nitrogen loss in soils as the ammonium was converted in nitrite, nitrate, and mainly in dinitrogen gas (Yang *et al.*, 2012). The exact microbe involved in this process is still unknown, but recent studies point towards *Acidimicrobiaceae* sp (Huang and Jaffé, 2015; Huang *et al.*, 2016). The oxidation of methane coupled to nitrate reduction to dinitrogen gas was first detected in a consortium of bacteria and

archaea (Raghoebarsing *et al.*, 2006). Further experiments revealed that a single bacterium, belonging to the NC10 phylum, was able to execute the process without the involvement of archaea (Ettwig *et al.*, 2008, 2009). In fact, this microorganism (*Candidatus* Methylomirabilis oxyfera) is able to produce oxygen through the dismutation of nitric oxide into oxygen and dinitrogen gas, and the intracellularly generated oxygen is used for the oxidation of methane (Ettwig *et al.*, 2010; Wu *et al.*, 2011). Little is known about the enzyme implicated in the nitric oxide dismutation (*aka* NOD) but it seems to be phylogenetically related to quinol-dependent NO reductases (qNORs) (Ettwig *et al.*, 2012).

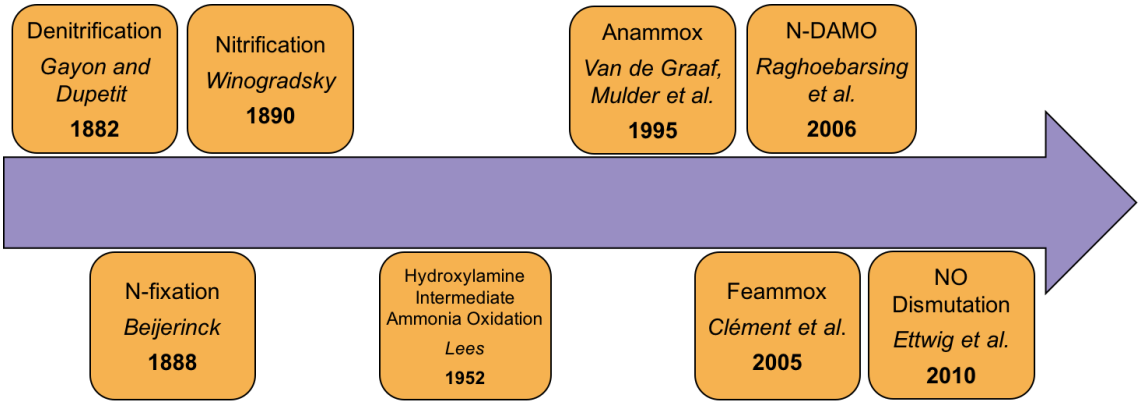

Figure 1.2. Historical timeline with new discoveries of processes related to the nitrogen cycle

Besides new discovered processes, also the knowledge of the microorganisms carrying out particular reactions within the nitrogen cycle has recently changed. After more than one hundred years associating ammonium oxidation to ammonium oxidizing bacteria (AOB), a marine archaeon was detected with the capacity to perform this process (Könneke *et al.*, 2005). Researchers have highlighted an important role of ammonia oxidizing archaea (AOA) in nitrification both in marine and terrestrial environments (Wuchter *et al.*, 2006; Francis *et al.*, 2007; Prosser and Nicol, 2008). A similar observation has occurred with the second step of the nitrification. For decades, *Nitrobacter* spp. were considered main responsible for nitrite oxidation. However, in the last years, other actors have been disclosed to be important in this process. In engineered systems, members of the genus *Nitrospira* were shown to be the dominant NOB (Schramm *et al.*, 1998; Juretschko *et al.*, 1998). On the other hand, *Nitrospina* spp. turned out to be the predominant marine NOB

reaching high abundances in oxygen minimum zones (Füssel *et al.*, 2012; Beman *et al.*, 2013). Additionally, other newly detected NOBs as *Nitrotoga* spp. could play important roles in some scenarios (Lücker *et al.*, 2014; Kinnunen *et al.*, 2017).

1.2 Nitrogen in natural and engineered systems

In the last hundred years, anthropogenic activities have greatly altered the biogeochemical nitrogen cycle on earth. The intense ammonium production for agricultural use as fertilizer via the Haber-Bosch process, together with the combustion of fossil fuels, entail that, nowadays, around half of the reactive nitrogen input to the biosphere comes from human activities. (Galloway and Cowling, 2002). In terrestrial environments, more nitrogen is fixed from anthropogenic activities than from natural sources (140 Tg N $year^{-1}$ and 110 Tg N $year^{-1}$, respectively) (Canfield *et al.*, 2010). A big part of this N is lost due to gaseous N-emissions and in the form of nitrate leaching out of soils (Cassman *et al.*, 2002). The reactive nitrogen, in form of N_2O, and ground-level ozone (created by reactions between NO_x and organic compounds) in the air, modify the balance of greenhouse gases. In addition, NO_x enhances the formation of secondary particulate matter that together with ground-level ozone cause respiratory diseases in humans. Air pollution also affects terrestrial ecosystems increasing soil acidification and deteriorating biodiversity (Erisman *et al.*, 2008; Sutton *et al.*, 2011). In water bodies, excess nitrate leads to eutrophication that provokes algal blooms impacting water life; and reduces the water quality both in surface and ground waters (Galloway *et al.*, 2003; Duce *et al.*, 2008; Erisman *et al.*, 2013).

To protect the environment from residual water discharges with high concentrations of ammonium and nitrate, biological nitrogen removal has been largely applied at wastewater treatment plants (WWTP). Conventionally, the approach adopted was based on microbial activity via autotrophic nitrification and heterotrophic denitrification. This process is costly because oxidation of ammonium requires constant supply of oxygen. With the discovery of new processes, several biological nitrogen removal treatments have been developed in order to improve the process performance and reduce the costs. One of the new processes most broadly studied are anammox-based technologies. This procedure reduces the oxygen demand up to 60% as well as eliminates the organic carbon needed for denitrification. Thus, the organic matter could be alternatively used to produce methane and subsequently energy via anaerobic digestion (Kartal *et al.*, 2010). However, there are some challenges

associated to anammox-based treatments which have to be further investigated as long start-up periods and poor effluent water quality (Ali and Okabe, 2015).

Besides treatment of residual water with high reactive nitrogen concentrations, waters with lower nitrogen concentrations may still require treatment to protect human health when water bodies are used for drinking water. Incomplete ammonium removal from drinking water treatments can lead to microbial regrowth and accumulation of nitrite in distribution systems (Chu *et al.*, 2005; Lytle *et al.*, 2007). Also, nitrite and nitrate have been associated with human health problems (Villanueva *et al.*, 2014; Jones *et al.*, 2016; Espejo-Herrera *et al.*, 2016). The most common sources for drinking water production are surface and groundwater. While reactive nitrogen may be present in both sources, surface water is typically characterized by a higher organic matter content, while groundwater usually has a higher mineral content and very low or no oxygen. The traditional approach for drinking water treatment consists of coagulation, flocculation, sedimentation, filtration and disinfection. In Denmark, where this study was conducted, groundwater is primarily used for drinking water production, and the main compounds in the water are NH_4^+, Fe^{+2}, Mn^{+2}, CH_4 and H_2S, all at or below the mg/L concentrations. As the aquifers have a relatively high water quality, few steps are required, and the main technology applied is biological filtration through granular media.

1.3 Biological filtration

Biological filtration is a commonly used drinking water technology consisting of passing the source water through a bed of granular filter material. There are several types of biological filters but the most commonly used in drinking water production are slow sand filters (SSF) and rapid sand filters (RSF). The main differences reside in the hydraulic loading rate and the filter cycle length, but other dissimilarities also contrast these biological filters (Table 1.1). Although SSF are simpler to operate, higher removal efficiencies have been observed in RSF (Crittenden *et al.*, 2005). Besides sand, other often utilized filter materials are anthracite, granular activated carbon (GAC) or dual media, which combine two of the mentioned materials. A deep investigation of different filter materials showed that GAC had a slightly better performance than the other materials, probably due to its irregular shape and macroporous structure, which offer larger spaces for bacteria attachment and better adsorption capacity (Urfer *et al.*, 1997). RSF has been widely used in Denmark showing an effective performance. The common steps followed in

the groundwater treatment are aeration, primary filtration and secondary filtration. According to textbook knowledge, the following happens in the treatment train (Gydesen and Tkker, 2013): The oxygenation favours stripping of CH_4 and H_2S. Then, water flows into the filters, typically by gravity, where oxidation of iron and manganese results in precipitates in form of oxyhydroxydes. In the secondary filters, the main processes are nitrification together with the oxidation of potential small amounts of remaining Fe^{+2}, Mn^{+2}, CH_4 and H_2S.

Table 1.1 Usual ranges of design and operational parameters in slow and rapid sand filtration (Tatari, 2014)[a].

Process characteristic	Slow sand filtration	Rapid sand filtration
Hydraulic loading rate (m/h)	0.05-0.20	3.0-15
Filter media diameter (mm)	0.30-0.45	0.50-1.2
Bed depth (m)	0.9-1.5	0.6-1.8
Required head (m)	0.9-1.5	1.8-3.0
Filtering cycle length	1-6 months	1-20 days
Ripening period	Several days	15 min- 2 h
Cleaning method	Scraping	Backwashing

[a]Values represent typical ranges and some filters are designed and operated outside these ranges

Although microbial activity has been suggested to play an important role in RSF, the actual microbial composition and the role of the microbial communities in the removal of primary pollutants have been poorly documented. Central metabolic functions have frequently been assigned to taxa based on their activity in other environments. Thus, *Nitrosomonas*, *Nitrospira*, *Gallionela*, *Bacillus* and *Methylococcaceae* spp. have been proposed to catalyze the removal of ammonium, nitrite, iron, manganese and methane, respectively in RSFs (Lautenschlager *et al.*, 2014; LaPara *et al.*, 2015; Vet *et al.*, 2012; Cerrato *et al.*, 2010; Albers *et al.*, 2015).

In a previous study, employing 16S rRNA amplicon sequencing, Proteobacteria and Nitrospirae were reported as the most abundant phyla in rapid sand filters at several Danish waterworks (Figure 1.3) (Gülay *et al.*, 2016). The dominance of *Nitrospira* spp. in the communities was especially surprising because members of this genus were known as nitrite oxidizers, and their measured abundances were much higher than those of known ammonium oxidizers. This might suggest that they have a role beyond nitrite oxidation in these systems. In addition, most of the other dominant taxa are poorly described with respect to physiology. Hence, it is not clear what physiological processes are truly mediated by many of the RSF community members.

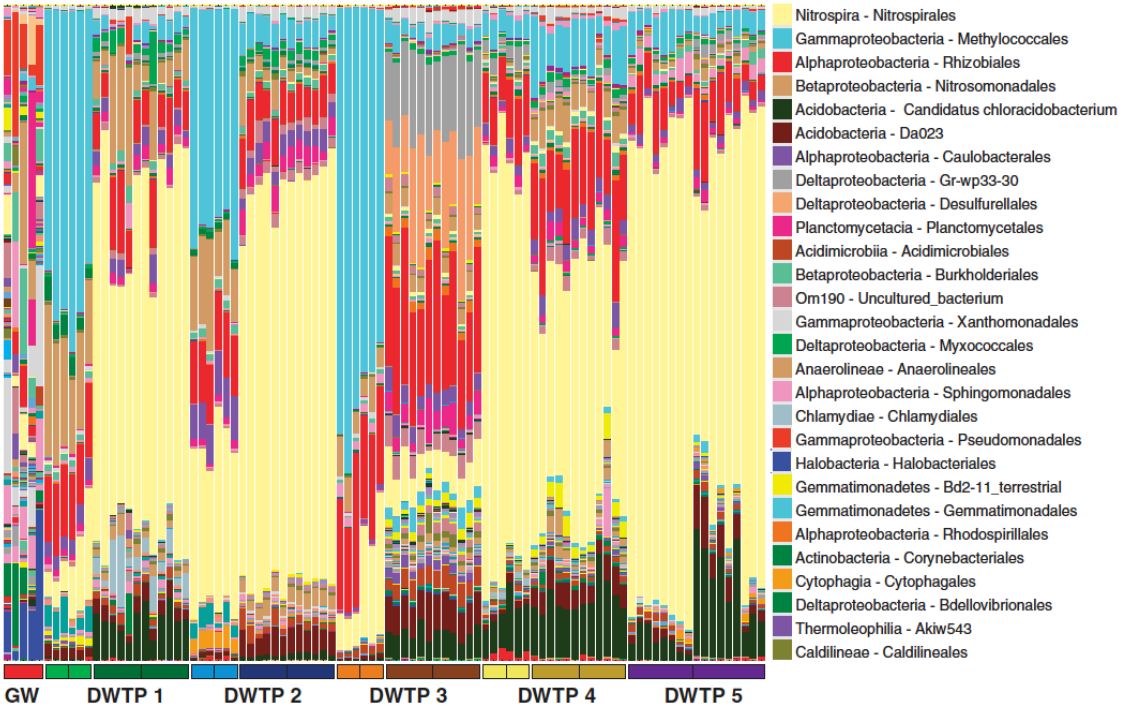

Figure 1.3. Relative abundance and order-level taxonomic classification of 16S rRNA amplicons across five drinking water treatment plants (Gülay *et al.*, 2016).

1.4 Research objectives and Thesis structure

The overall aim of this PhD project was to unravel the metabolic potential of microbial communities in groundwater-fed biological filters, identify the role of the dominant actors and elucidate the reasons behind the previously observed unexpected high abundance of *Nitrospira* spp. For this purpose, metagenomics analysis as well as in situ measurements were carried out both in full-scale and with lab-scale biofilters. The detailed main aims of the PhD were:

- Comprehensively investigate the microbial community of a rapid sand filter via metagenomics analysis and genome binning. (Paper I)
- Describe the genomic capabilities of comammox *Nitrospira*. Perform a comparative genomics analysis of the available comammox and canonical *Nitrospira* genomes. Shed light into the ecological niche and evolution history of comammox *Nitrospira* (Paper II)
- Quantify the abundance and diversity of nitrifying communities in groundwater-fed biological filters at different waterworks (Paper III)
- Reveal the main factors driving the abundance of *Nitrospira* over other nitrifiers in drinking water biofilters using a lab-scale experimentation (Paper IV)

Chapter 2 presents an overview of techniques used in microbiology from the initial cultivate-dependent methods to newly next-generation sequencing-based approaches.

Chapter 3 describes the genomic capacity of the recover population genomes form a rapid sand filter and their potential interaction with the constituents present in the groundwater. Additionally, it also contains the evidences of the discovery of a composite population genome with the capability for complete ammonium oxidation.

Chapter 4 details the genomic capacity of individual comammox *Nitrospira* population genomes. This chapter also contains a comparative genomic analysis of the comammox genomes with canonical *Nitrospira* and AOB genomes. Moreover, it focuses on the evolutionary history of comammox *Nitrospira*.

Chapter 5 introduces the new primers designed to exclusively target comammox *Nitrospira* and evaluates the abundance of comammox *Nitrospira* and other nitrifiers in several rapid sand filters.

Chapter 6 gives an overview of technologies used to link microbial activity with the phylogenetic identities of uncultivated microorganisms. This chapter also includes the description of a lab-scale experiment with filter material containing different relative proportions of comammox and canonical *Nitrospira* as well as AOB and AOA, exposed to different ammonium loading conditions.

2 From amplicon sequencing to metagenomics

For many years, environmental microbiologists have tried to understand which microorganisms are responsible for the main biogeochemical processes occurring in ecosystems. The traditional approach has consisted in collecting samples from the environment and cultivating them in laboratory-prepared media either trying to mimic in situ conditions or enriching microbes carrying out specific metabolic reactions by incubating with particular compounds (Madsen, 2005). Although this method allowed to isolate and identify microorganisms involved in different biogeochemical processes, several limitations have been observed. Frequently, just a minor fraction of the community can be cultivated and the isolated microorganisms not always reflect their relative abundance under natural conditions (Amann *et al.*, 1995). The difficulty to accurately reproduce native conditions, and the complex and dependent microbial interaction ongoing in the natural environments are some of the reasons that limit cultivation-based approaches. However, new high-throughputs cultivation methods such as those that involve diffusion chambers and isolation chips (Kaeberlein, 2002; Nichols *et al.*, 2010), hollow-fiber membrane chambers (Aoi *et al.*, 2009) or gel microdroplets (Zengler *et al.*, 2002) are being developed to try to circumvent some of these limitations.

2.1 16S rRNA gene sequencing

In the last two decades, cultivation-independent methods have notably improved our knowledge about microbial communities in natural and engineering systems overcoming cultivation-related limitations.

Particularly, methods based on targeting the small subunit of the rRNA gene have been popular. This approach allows to taxonomically and phylogenetically characterize microbial communities as well as their structure and dynamics. This is possible due to the universal presence across all domains of life of the small subunit rRNA gene (16S rRNA in prokaryotes and 18S rRNA in eukaryotes), their relatively homogeneous evolution rate, and the presence of conserved and hypervariable regions in this gene. This technique has facilitated the discovery and classification of an extensive diversity of uncultivated microorganisms (Tringe and Hugenholtz, 2008). Different techniques based on the 16S rRNA gene have been developed such as cloning libraries, fingerprinting, fluorescence in situ hybridization (FISH) or most

recently amplicon sequencing. With time, FISH for visualization of spatial distribution and structure of microbial communities, and amplicon sequencing for community identification and diversity calculation have become dominant techniques in microbial ecology. The main advantages of the 16S rRNA amplicon sequencing approach to describe microbial community composition are the cost-to-output ratio, the well-established pipelines for its analysis and the vast public databases for references, which makes this approach attractive for studies with complex communities and large number of samples. However, its limitations include occurrence of sequencing errors, PCR amplification biases or output differences depending of the chosen gene region (Zhou *et al.*, 2015). In addition, 16S rRNA gene sequencing approaches only predict phylogeny but most of times no information is directly gained about physiology.

In the last years, tools have been developed to predict functional capacity using 16S rRNA gene such as PICRUSt (Langille *et al.*, 2013) and Tax4Fun (Aßhauer *et al.*, 2015). Although they have successfully captured key functions of well-studied microbial communities mainly related to the human microbiome, they are limited in application for poorly investigated environmental samples (Iwai *et al.*, 2016) due to their dependency on fully-sequenced reference genomes.

Another approach to try to uncover the physiological capacity of a microbial community is the use of specific genes markers (functional genes) targeting particular functions. For instance, the ammonium monooxygenase subunit A (*amoA*) and the nitrogenase reductase (*nifH*) have been widely used as marker genes to detect the first step of nitrification and N-fixation, respectively (Levy-Booth *et al.*, 2014). The main limitation of this technique is that the primers used to target specific functions typically cannot capture all the microorganisms which can carry out the studied physiology (Palmer *et al.*, 2012; Penton *et al.*, 2013; Dechesne *et al.*, 2016).

2.2 Shotgun metagenomics

Shotgun metagenomics sequencing bypasses most of methodological issues described for amplicon sequencing and in addition allows not only taxonomical identification, but also discovery of physiological potential. In this approach, DNA is extracted from an environmental sample and sheared into small fragments which are all sequenced without targeting a specific gene for amplification. The first generation of sequence technologies, dominated by the Sanger sequencing, were characterized by high accuracy and medium DNA sequence (called read) size. In the second generation, there was gener-

ally a decrease in the read size, a maintained accuracy, and a highly increased total sequencing output size per run, especially with the Illumina technology. Lastly, in the third generation, the main objective has been to increase the read size although so far, both accuracy and the total bases sequenced per run has decreased (Table 2.1).

Table 2.1. Overview of sequencing technologies (Adapted from Ghurye *et al.*, 2016)

Technology	Read Length	Accuracy	Time per run	Bases per run	Generation
Sanger sequencing	400 to 900 bp	99.9% (High)	20 min. to 3 h	50 – 100 Kb	First
Pyrosequencing (454)	700 bp	98% (Medium)	24 hours	400 Mb	Second
Illumina	50-300bp	99.9% (High)	1 to 11 days	300 Gb	Second
SOLiD	75 bp	99.9% (High)	1 to 2 weeks	3 Gb	Second
Ion Torrent	Up to 400 bp	98% (Medium)	2 hours	10Gb	Second
Pacific Biosciences	10 kbp to 15 kbp	87% (Low)	30 min. to 4 h	5 – 10 Gb	Third
Oxford Nanopore MinION	5 kbp to 10 kbp	70% to 90% (Low)	1 to 2 days	500 Mb	Third

Reads are first subject to a quality control followed by a trimming of the low quality bases and the adapters used during the DNA preparation before sequencing. FastQC (Andrews, 2010) is one of the most used tools for quality control, while Trimmomatic (Bolger *et al.*, 2014) and cutadapt (Martin, 2011) are frequently used for trimming. In a next step, surviving reads are typically either mapped against reference sequences or genomes, or pieced together (forming contigs) using assembly algorithms (*de novo* assembly). Read mapping is based on algorithms which align the sequences against reference genomes to identify regions of similarity. A recent study showed that BWA (Li and Durbin, 2010) (used in **Paper I**) and Bowtie2 (Langmead and Salzberg, 2012) are the most accurate mapping tools (Ziemann, 2016). Read alignment is restricted by existence of reference genomes, as well as other issues related to polymorphic regions or large insertions. As for *de novo* assembly, the majority of the modern assembler relies on the de Bruijn graph approach in which reads are split into short fixed length sequences referred to as k-mers (k is the length of the short sequence). The graph is constructed using the k - 1 prefix and suffix of the k-mers as nodes connected by edges that represent the k-mers. Then, the constructed graph is condensed, the optimal path is selected and transformed into a contiguous sequence (contig) (Compeau *et al.*,

2011; Ghurye *et al.*, 2016). In a recent evaluation of different assembly tools for metagenomics analysis, metaSPAdes (Nurk *et al.*, 2017), IDBA-UD (Peng *et al.*, 2012) (used in **Paper I** and **II**) and Megahit (Li *et al.*, 2014) (used in **Paper II**) showed the best assembly performance (Vollmers *et al.*, 2017). Gene prediction is then executing on the assembled contigs to find coding sequences (CDS). Gene finding softwares are usually based on machine learning algorithms (Prodigal (Hyatt *et al.*, 2010); used in **Paper I** and **II**), codon usage combined with Hidden Markov Model (HMM) (MetaGeneMark (Zhu *et al.*, 2010)) or sequencing error model and codon usage combined with HMM (FragGeneScan (Rho *et al.*, 2010)) (Kim *et al.*, 2013). Predicted CDS are generally both taxonomically and functionally annotated.

Taxonomical annotation can be obtained by searching homology between the predicted CDS and a combination of conserved marker genes. This approach is applied in tools as PhyloSift (Darling *et al.*, 2014) (used in **Paper I**) or MetaPhlAn (Segata *et al.*, 2012), which also includes unique clade-specific marker genes. In turn, other software as MEGAN (Huson and Weber, 2013) (used in **Paper I** and **II**) follows a lowest common ancestor (LCA) approach on the blast hits obtained by a similarity search. In addition, other tools assign taxonomy directly taking assembly-free reads. These algorithms typically rely on mapping reads against reference sequence databases (MGmapper (Petersen *et al.*, 2017)), k-mer comparison with reference genomes (Kraken (Wood *et al.*, 2014)) or in sequence comparison to a reference database of microbial proteins (Kaiju (Menzel and Krogh, 2015)).

Functional annotation provides a characterization of the genomic capacity of the investigated microbial community. The most common approach is based on searching homologies (generally using BLAST) of the predicted CDS against a database. Besides broad protein sequences databases as NCBI-NR and UniProt, other databases include functional categories (COG (Tatusov, 1997) and eggNOG (Powell *et al.*, 2012)), metabolic pathways (KEGG (Ogata *et al.*, 1999)), subsystems (SEED (Overbeek *et al.*, 2005)) or protein and domain families (Pfam (Punta *et al.*, 2012) and TIGRFAM (Haft, 2003)). In the last case, the search is usually carried out using the hidden Markov model-based algorithm (HMMER (Finn *et al.*, 2011)).

2.3 Recovery of population genomes

Recently, a further step has been taken to better understand the role of the main types in microbial communities; this involves genome recovery from metagenomes. In this case, prior to gene prediction, the assembled contigs are clustered into separate taxonomic groups. This process is referred to as binning and can be performed based on sequence composition, differential abundance or a combination of both (Figure 2.1). The first approach relies on the

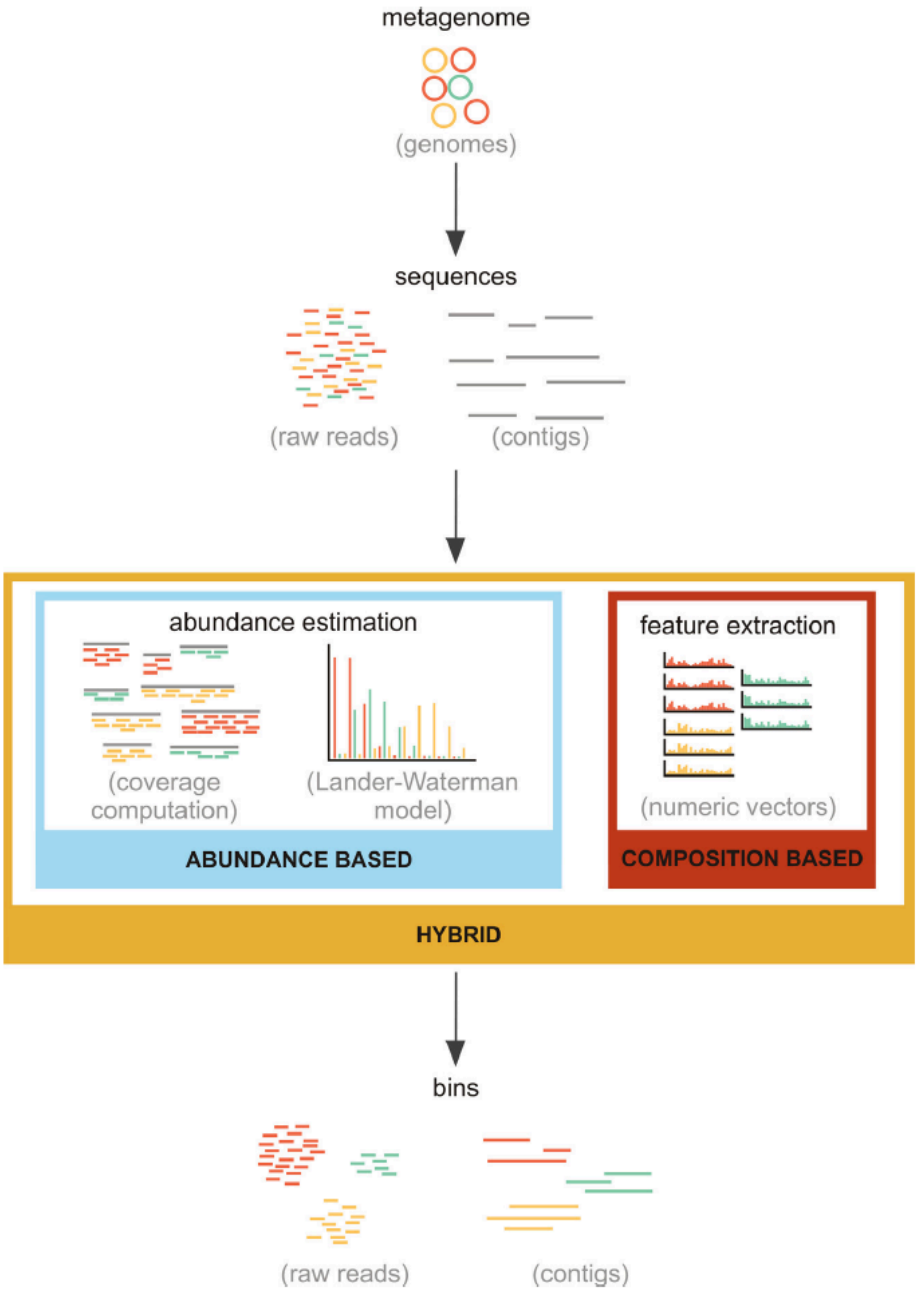

Figure 2.1. Workflow of binning strategies (Sedlar *et al.*, 2017)

premise that the genomic composition is exclusive for each species. This observation was first done looking at the %GC of multiple genomes (Karlin *et al.*, 1997): the %GC clearly changes between species but is relatively constant within species. Further studies showed an even more effective discrimination based on oligonucleotide frequencies (typically 4 or 5-mers) (Sandberg *et al.*, 2003; Pride, 2003). Although the exactly reasons behind this observation remain unclear different studies have pointed towards neutral and selective processes such as the replication, repair, and recombination cell machinery or external factors (as temperature, pH or niche complexity) (Foerstner *et al.*, 2005; Dick *et al.*, 2009). Accordingly, it is possible to group contigs by comparing their genomic signatures. On the other hand, the differential abundance method is subject to the assumption that abundance profiles of contigs from the same genome are correlated across different samples where that particular genome is present. In the last year, multiple binning algorithms have been released. For instance, Metawatt (Strous *et al.*, 2012) and VizBin (Laczny *et al.*, 2015) (used in **Paper I**) follow a genomic signature-based strategy; Canopy (Nielsen *et al.*, 2014) relies on the different abundance across multiple metagenomics samples, and CONCOCT (Alneberg *et al.*, 2014), MetaBat (Kang *et al.*, 2015) or mmgenome (Karst *et al.*, 2016) (used in **Paper II**) combine both sequence and abundance information.

Metagenomic analysis and recovery of population genomes have enabled the discovery of a variety of novel microbial pathways and new microorganisms both in natural and engineering systems (Wexler *et al.*, 2005; Bryant *et al.*, 2007; Carrión *et al.*, 2015; Nobu *et al.*, 2015; Brown *et al.*, 2015).

In this PhD project, the central objective was to unravel the physiology of the main types of the microbial community in groundwater-fed rapid sand filters and relate it with the environmental conditions in the biofilter. Hence, I performed a metagenomic analysis together with a binning approach based on sequence composition to recover the dominant populations of the mentioned community (**Paper I**). In addition, as I hypothesized that the high abundance of *Nitrospira* spp. could be explained by a possible extra physiology beyond their nitrite oxidation capacity, techniques based on 16S rRNA or in already sequenced reference genomes would not decipher new potential genetic capacities. Thus, an abundance-based approach was executed to individually separate a *Nitrospira* spp. composite population genome followed by a functional annotation to exhaustively analyse the genome's genetic capabilities (**Paper II**).

3 Discovery of comammox

3.1 Nitrifying microorganisms

Since the investigations by Winogradsky more than 100 years ago (Winogradsky, 1892), nitrification had been considered a two-step process: oxidation of ammonium to nitrite by ammonium-oxidizing prokaryotes (AOP) and oxidation of nitrite to nitrate by nitrite-oxidizing bacteria (NOB). AOP are phylogenetically restricted to the three bacterial genera: *Nitrosococcus* (Gammaproteobacteria), *Nitrosomonas* and *Nitrosospira* (both Betaproteobacteria); and members of the archaeal phylum Thaumarchaeota. On the other hand, broader diversity has been detected for NOB, three genera in the Proteobacteria phylum: *Nitrobacter, Nitrotoga* and *Nitrococcus* belonging to Aplha-, Beta- and Gammaproteobacteria, respectively; and four other genera assigned to *Nitrospira* in the Nitrospirae phylum, the marine NOB *Nitrospina* and the recently discovered *Candidatus* Nitromaritima, in the Nitrospinae phylum, and the only Gram-positive NOB, *Nitrolancea* in the phylum Chloroflexi.

Thermodynamic considerations show that the free energy available in the oxidation of ammonium to nitrite ($\Delta G^{\circ\prime} = -275$ kJ mol^{-1}) is higher than energy accessible from the oxidation of nitrite to nitrate ($\Delta G^{\circ\prime} = -76$ kJ mol^{-1}). This would imply that AOP should outnumber NOB in microbial communities. Such is consistent with a wide number of studies which followed this assumption (Ye *et al.*, 2011; Albertsen *et al.*, 2012; Rughöft *et al.*, 2016; Speth *et al.*, 2016). However in some cases, the abundance of detected NOB has been up to two orders of magnitude higher than the AOP abundance (Schramm *et al.*, 1999; Foesel *et al.*, 2008). This observation has been particularly frequent in drinking water biofilters and distribution systems (Martiny *et al.*, 2005; Feng *et al.*, 2012; White *et al.*, 2012; LaPara *et al.*, 2015; Albers *et al.*, 2015; Cai *et al.*, 2015; Nitzsche *et al.*, 2015). A possible explanation could be that the growth of NOB is not only supported by nitrite oxidation, but also by other substrates. In fact, some NOB can use other electron donors such as hydrogen, formate or simple organic compounds (Watson *et al.*, 1986; Gruber-Dorninger *et al.*, 2014; Koch *et al.*, 2014). Nevertheless, these compounds are not always present in the aforementioned systems and it could not explain the dominance of NOB in some of those environments. Another alternative explanation is the putative capacity of *Nitrospira* spp. to completely oxidize ammonium to nitrate. In 2006, Costa *et al*. proposed that such

organisms could exist (Costa *et al.*, 2006). They highlighted that full nitrification is energetically favourable and that, although -compared to canonical ammonium oxidizers- a complete ammonium oxidizer would have a lower specific growth rate, it would have higher yield per unit ammonium removed. Moreover, they predicted that this metabolism would be advantageous in aggregates or biofilms with low substrate availability.

In Paper I, the microbial community of a surface attached environment with low ammonium inputs where *Nitrospira* spp. highly outnumbered AOP (Gülay *et al.*, 2016), was investigated through metagenomics.

3.2 Main actors in a rapid sand filter

From the rapid sand filter metagenome, 14 near-complete population genomes were recovered following a sequence-composition based binning approach (Chapter 2). The reconstructed population genomes have the genetic capacity to grow on the typical compounds found in the source water (Figure 3.1). Thus, one population genome harbours the genes to oxidize methane.

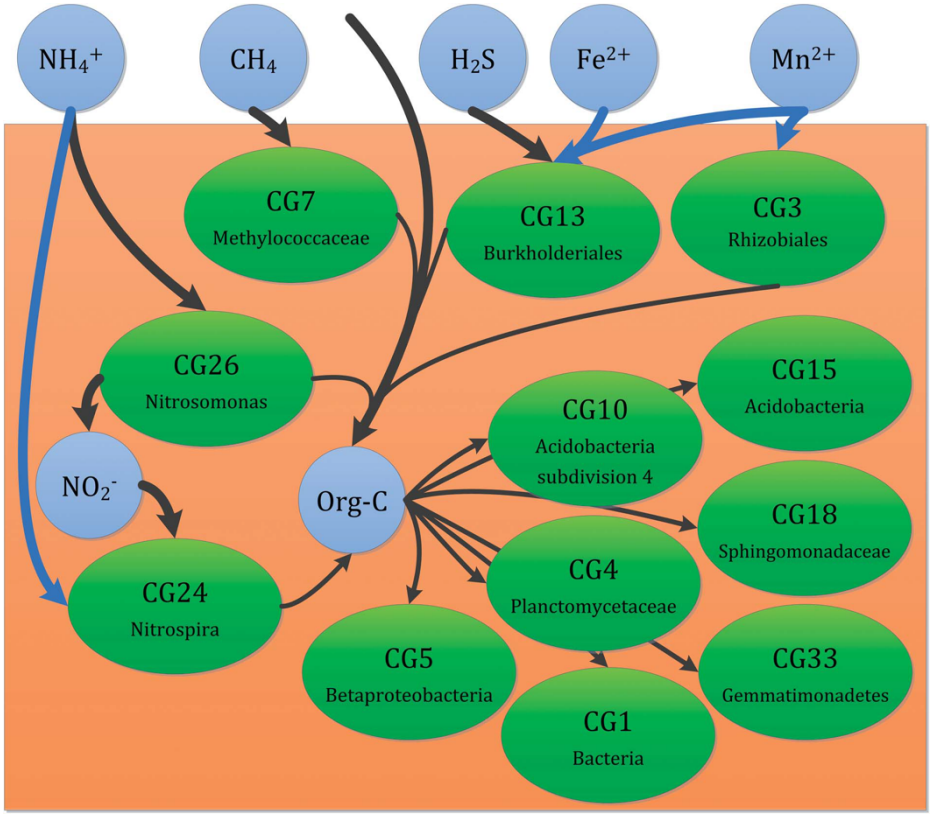

Figure 3.1. Model of predicted metabolic and geochemical processes facilitating the degradation of groundwater contaminants in rapid gravity sand filters based on metagenomic analysis. Grey arrows denote metabolic capability, whereas blue arrows denote putative metabolic capability. (Paper I)

Another genome has a potential high versatility containing genes that could be involved in sulphide, iron and manganese oxidation. Manganese could potentially also be oxidized by a heterotrophic population genome. Several of the recovered population genomes encoded for pathways associated with the degradation of organic compounds. In relation to nitrification, a *Nitrosomonas*-related draft genome encodes the pathway for ammonium oxidation (Further details about the genomic capabilities of the population genomes can be found in **Paper I**) and sequences affiliated to AOA were also detected in the metagenome. As expected, a high abundance composite population genome was assigned to *Nitrospira* spp. This composite genome could not be separated into individual genomes. The difficulty to separate these genomes probably derives from micro-diversity, which is a hindrance to genome reconstruction and segregation from metagenomics data (Wilmes *et al.*, 2009). The *Nitrospira* genomic content investigation showed that besides the characteristics genes related to nitrite oxidation, this composite population genome has several contigs containing genes of the ammonium oxidation pathway. These genes (referred to as atypical-AMOX genes) are phylogenetically dissimilar from the ones found in canonical AOPs (here referred to as typical-AMOX genes), being most closely related to betaproteobacterial AOB counterpart genes (Figure 3.2).

3.3 Evidence of comammox *Nitrospira* detection

Considering the novelty of this finding, we exhaustively investigated the (meta)genomic data to rule out potential contamination during the binning process and provide strong evidence of a complete ammonium oxidation pathway in the recovered composite population genome. The first evidence is based on the observation that the GC content of a genome is usually uniform along their genetic content unless it has been recruited by recent horizontal gene transfer (Karlin *et al.*, 1997). A similar case is valid for the sequencing depth, as it is assumed that each nucleotide stretch coming from a specific genome would be sequenced a similar number of times (Glusman *et al.*, 2015)

Thus, the GC content and sequencing depth of the atypical-AMOX genes and typical-AMOX genes were compared with the average GC content and sequencing depth of the *Nitrospira* and *Nitrosomonas* population genomes. In both cases, the distribution of atypical-AMOX genes was identical to *Nitrospira* (identified as CG24) (Wilcoxon test, P-value>0.05) but different from

Nitrosomonas population genomes (identified as CG26) (Figure 3.3) (**Paper I)**.

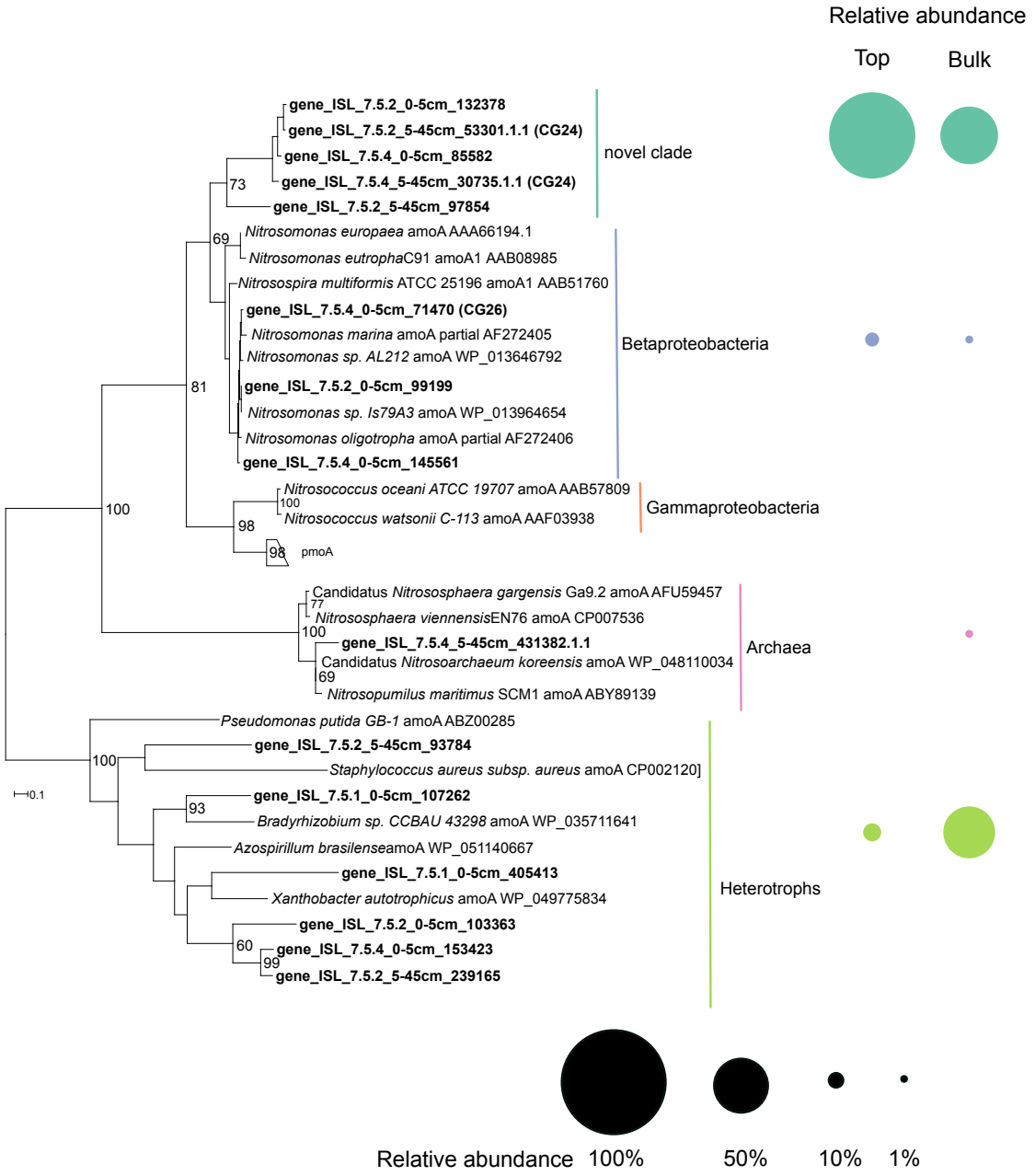

Figure 3.2. Phylogenetic reconstruction of *amoA* and *pmoA* reference protein sequences with putative *amoA* sequences recovered from the metagenomes (bold). Relative abundance of putative *amoA* clusters in the metagenomes is shown with colours corresponding to phylogenetic groups. Bootstrap support greater than 60 is indicated (based on 100 replicates). **(Paper I)**

The second line of evidence derives from the notion that abundance is constant across genetic entities (i.e., every gene on a specific genome will be detected in a sample in equal abundance as any other gene from that genome) (Nielsen *et al.*, 2014). Based on this assumption, we investigated, the covariance between the abundance of genes of interest in the whole metagenome

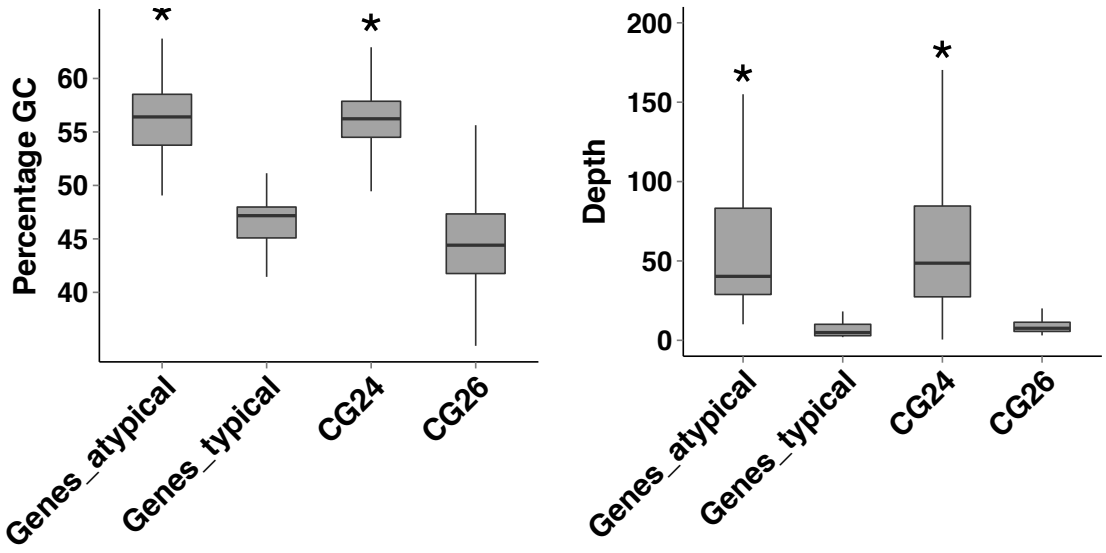

Figure 3.3. Evidences of presence of AMOX related genes in Nitrospira CG24. a) Comparison of GC content in atypical AMOX genes, typical AMOX genes, *Nitrospira* CG24 and *Nitrosomonas* CG26. b) Comparison of depth in atypical AMOX genes, typical AMOX genes, *Nitrospira* CG24 and *Nitrosomonas* CG26. **(Paper I)**

and the investigated population genomes (CG24 and CG26) along six samples. All atypical-AMOX genes statistically correlated with *Nitrospira* CG24 ($R2>0.97–0.99$, P-value<0.001) but not with *Nitrosomonas* CG26 ($R2>0.24–0.51$, P-value>0.1; Figure 3.4) (**Paper I)**. Taken together, these results strongly support the finding that the *Nitrospira* composite genome CG24 contains genes encoding for ammonia oxidation.

Additionally, based on the presence of essential single copy genes (average of 4.7 single-copy genes) and the number of *amo* operons (three) in the *Nitrospira* composite population genome, we estimated that three out of the five potential individual *Nitrospira* genomes are complete ammonium oxidizers.

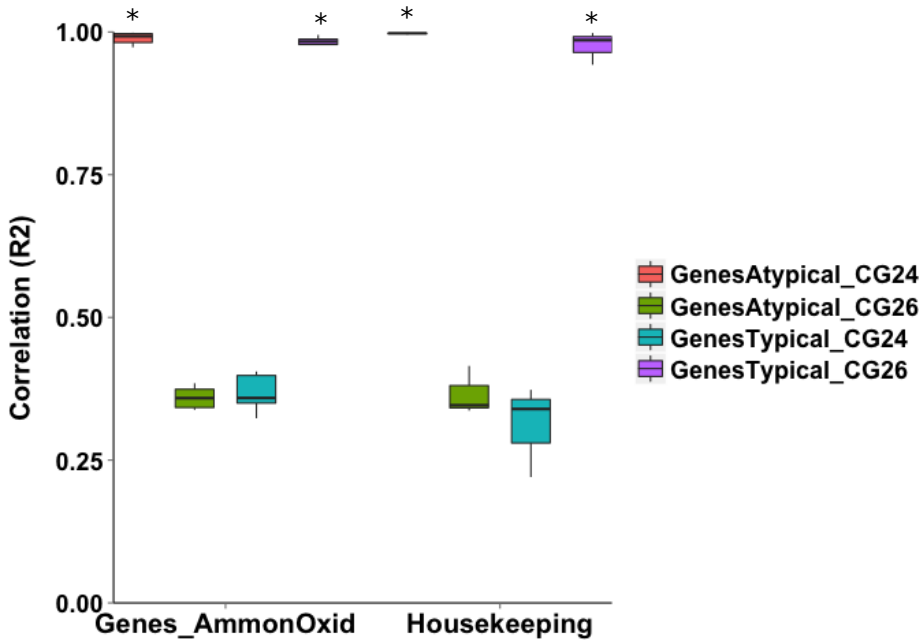

Figure 3.4. Correlation (R2) of abundance across six RSF samples of atypical AMOX genes (GenesAtypical in the legend) and typical AMOX genes (GenesTypical in the legend) in the gene catalog and *Nitrospira* CG24 and *Nitrosomonas* CG26 (panel left). Correlation (R2) of abundance across six RSF samples of *Nitrospira* housekeeping genes (GenesAtypical in the legend) and *Nitrosomonas* housekeeping genes (GenesTypical in the legend) in the gene catalog and Nitrospira CG24 and *Nitrosomonas* CG26 (panel right). **(Paper I)**

3.4 Simultaneous discovery of comammox in other environments

At a similar time of the detection of comammox *Nitrospira* in a rapid sand filter, three other groups identified single microorganism or genomes with the capacity to carry out complete ammonium oxidation. Daims *et al*. (2015) enriched a culture obtained from a biofilm developed on the walls of an oil exploration well. This enrichment was able to oxidize ammonium to nitrite with CO_2 as the sole carbon source and ammonium as the only energy source. The reconstruction of genomes from DNA extracted from the enrichment showed a *Nitrospira* sp. containing the pathways for complete ammonium oxidation. The other recovered genome from the enrichment was not associated with ammonium oxidation and it did not harbour any of the genes related to nitrification. Additionally, they observed expression of the enzymes involved in ammonium and nitrite oxidation through metaproteomic analysis. The comammox organism was named *Candidatus* Nitrospira inopinata. Similarly,

van Kessel *et al*. (2015) enriched a community from an anaerobic compartment of a trickling filter connected to a recirculation aquaculture system. The enrichment community was mainly composed of an anaerobic ammonium oxidizer and *Nitrospira* spp. DNA sequencing and genome reconstruction revealed two *Nitrospira* genomes containing the genes involved in ammonium and nitrite oxidation. In addition, they used a fluorescently labelled ammonia analogue that binds to the ammonia monooxygenase (AMO) with *Nitrospira*-specific fluorescence *in situ* hybridization (FISH) probes, disclosing the expression of a functional AMO by these *Nitrospira* spp. They showed that both *Nitrospira* types are capable of CO_2 fixation coupled to ammonia oxidation combining FISH and ^{14}C-labelling. They named these two genomes as *Ca.* N. nitrosa and *Ca.* N. nitrificans. Lastly, Pinto *et al*, (2015) recovered a draft genome from a drinking water system metagenome, encoding the pathway required for complete ammonium oxidation.

These joint discoveries support another dramatic revision of the nitrogen cycle (Figure 3.5).

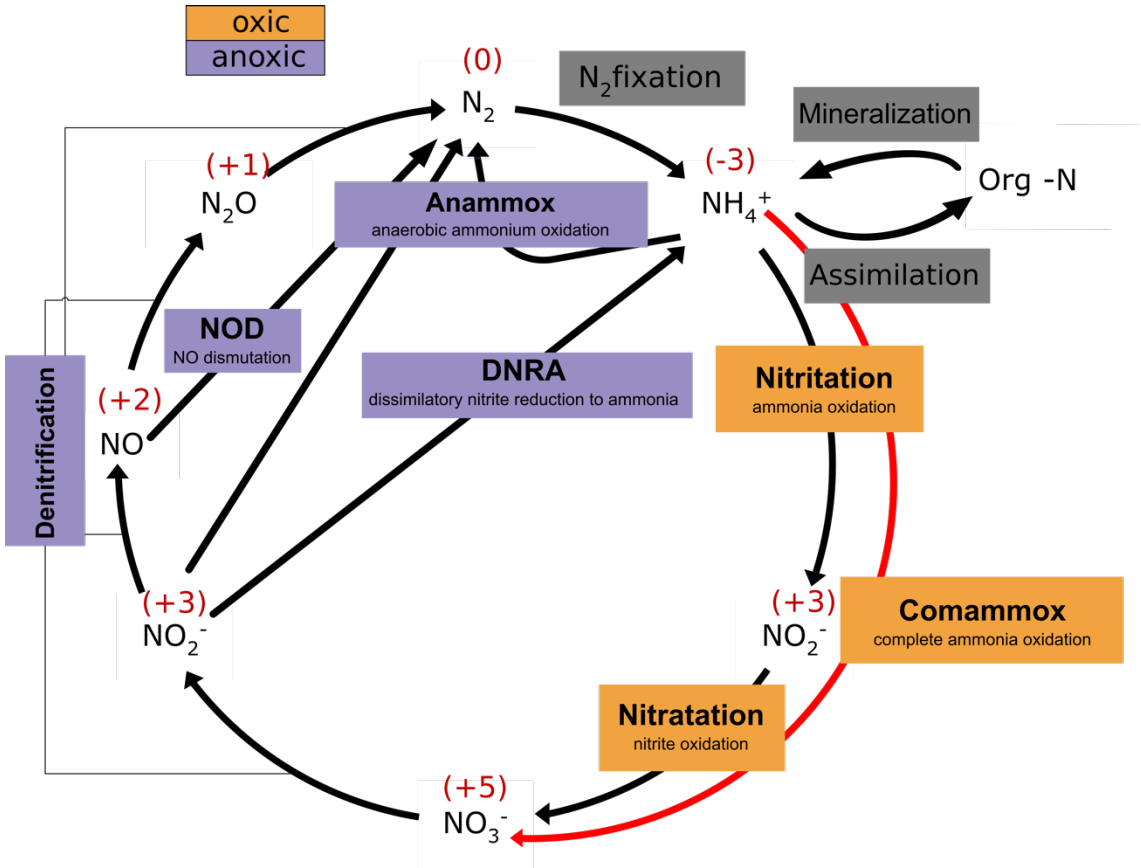

Figure 3.5. Major processes of the nitrogen cycle including the complete ammonia oxidation (comammox).

4 Description of comammox

The discovery of complete ammonium oxidizing microorganisms has opened several questions related to their ecology, physiology and evolution. Investigation of the genomic content of the available comammox genomes might help to shed light on these inquires.

Genomes with the capacity for complete ammonium oxidation are affiliated to the *Nitrospira* genus which is a highly diverse monophyletic group divided in at least six sublineages based on 16S rRNA similarity (Lebedeva *et al.*, 2011). A crucial enzyme of *Nitrospira* as well as of other NOB is the nitrite oxidoreductase (NXR), which catalyses the oxidation of nitrite to nitrate. This protein belongs to the dimethyl sulphoxide reductase family and specifically to the type II group of molybdopterin-binding enzymes (Meincke *et al.*, 1992). NXR is a membrane-associated enzyme with the catalytic site localized either in the cytoplasm (*Nitrobacter*, *Nitrococcus*, and *Nitrolancea*) or in the periplasmic space as in the case of *Nitrospira* (as well as *Nitrospina* and *Candidatus* Nitromaritima) (Daims *et al.*, 2016). NXR consists of a catalytic α subunit (NxrA), a γ subunit (NxrC) likely acting as the membrane anchor and a β subunit (NxrB) which channels electrons derived from nitrite to the electron transport chain (Lucker *et al.*, 2010).

A first general overview indicates that the nitrite oxidoreductases (NXR) of comammox *Nitrospira* genomes are highly similar to the NXRs of strictly nitrite-oxidizing *Nitrospira* (Daims *et al.*, 2015; van Kessel *et al.*, 2015; **Paper II**). Based on these features, it could be thought that comammox *Nitrospira* would have similar genomic characteristics as canonical *Nitrospira* besides the capacity to carry out the first step of nitrification.

4.1 Recovery of individual *Nitrospira* genomes

To further inspect the genomic composition of comammox *Nitrospira* and find similarities and distinctions between comammox and canonical *Nitrospira* and AOB, as well as within comammox clades, an effort was applied to separate the *Nitrospira* composite genome (CG24) recovered from the rapid sand filter community (**Paper I**) into individual genomes. As the sequence composition-based binning approach was not able to separate CG24, we conducted a differential abundance binning approach (Chapter II) using six different community samples collected from the same RSF. With this method (for further details SI of **Paper II**), we were able to extract five *Nitrospira* population genomes; CG24_A, CG24_B, CG24_C, CG24_D and CG24_E.

Genes required for complete ammonium oxidation (*amo* and *hao* operons) were detected in four of the recovered genomes. The protein sequence comparison of the AmoA revealed that CG24_A, CG24_C and CG24_E clustered in the same group and had relatively high sequence dissimilarity with CG24_B. This AmoA sequences divergence in comammox genomes (as well as for the other AMO subunits and for HAO) was also observed by others (Daims *et al.*, 2015), and based on that, two comammox clades (A and B) have been proposed. Thus, CG24_A, CG24_C and CG24_E belong to the clade B while CG24_B affiliates with clade A (Figure 4.1).

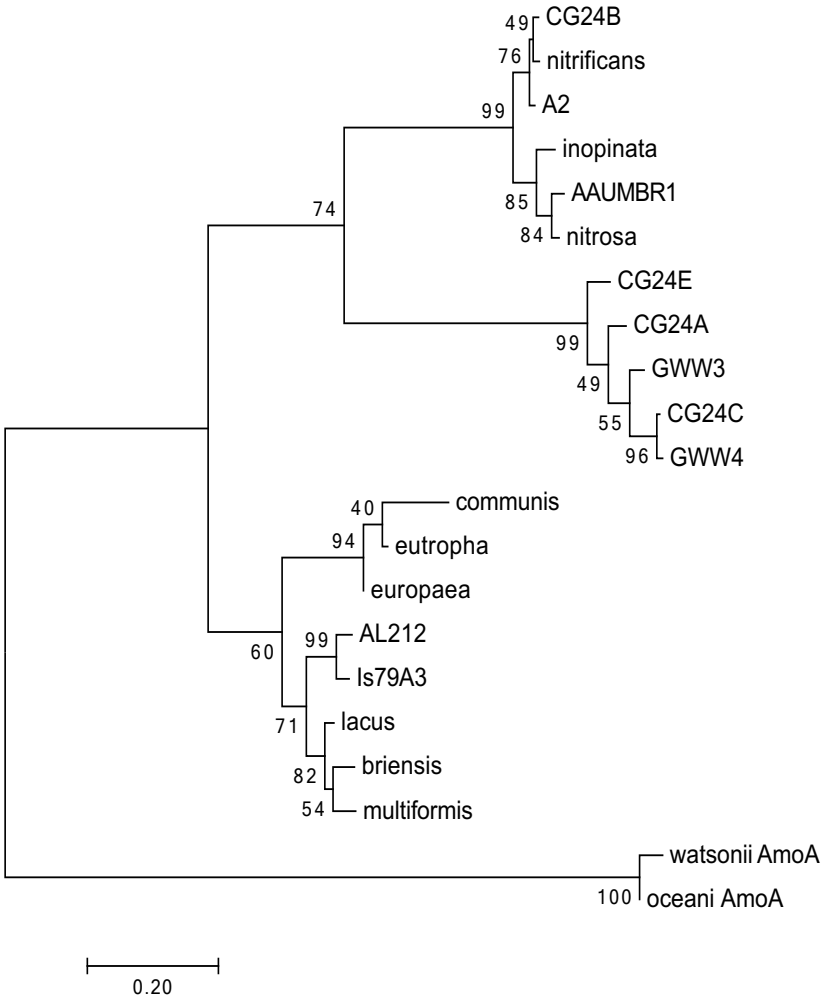

Figure 4.1. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.9313)). There are a total of 247 positions in the final dataset. **(Paper II)**

A concatenated analysis of 14 ribosomal proteins showed that the five recruited genomes belong to the *Nitrospira* lineage II (SI in **Paper II**) as observed for all so far discovered comammox genomes.

4.2 Comparative analysis of *Nitrospira* genomes

To gain deeper insights into the genetic capacity of comammox *Nitrospira*, we conducted a comparative genomics analysis (**Paper II**) of 16 *Nitrospira* genomes (11 comammox and 5 canonical *Nitrospira)* including the here recovered genomes together with publicly-available, high quality (>90% completion, <5% redundancy) *Nitrospira* genomes (Table 1, **Paper II**). These 16 genomes could be separated into 11 different species (7 of them comammox *Nitrospira*) based on an average amino acid identity (AAI) analysis (species level of 85% AAI similarity (Luo *et al.*, 2014); Figure 4.2).

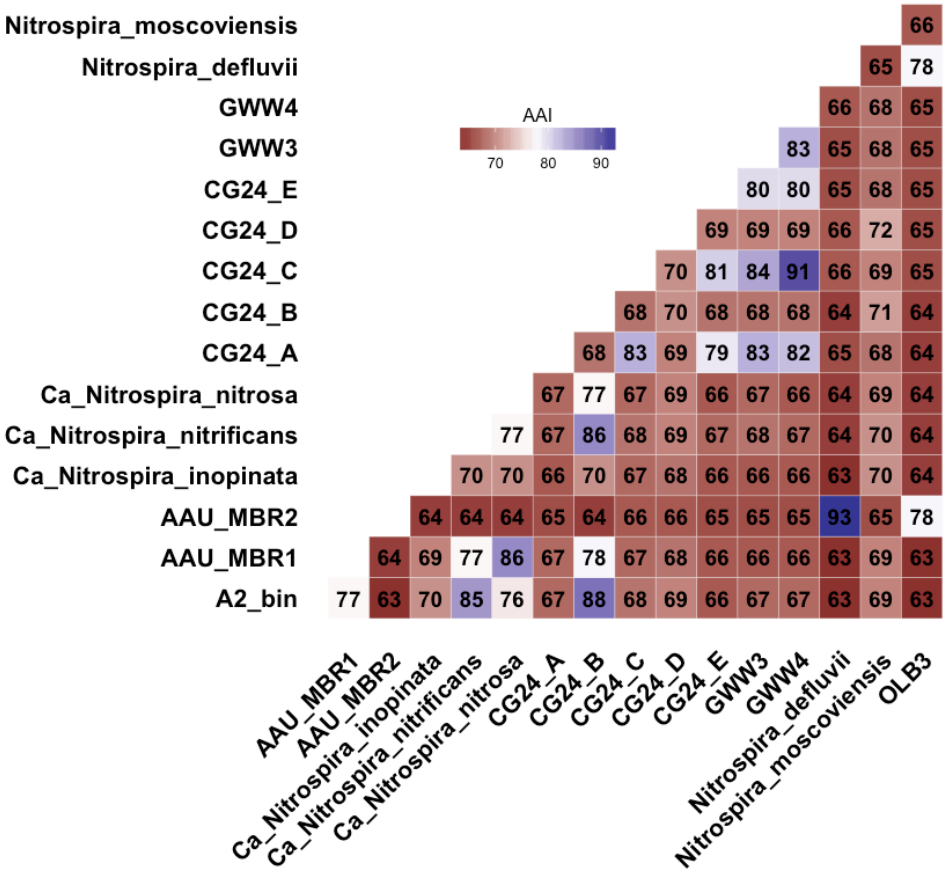

Figure 4.2. Pairwise average amino acid identity (AAI) comparison of the 16 *Nitrospira* genomes. **(Paper II)**

The coding sequences (CDS) of the 16 examined *Nitrospira* genomes accounted for 59,744 CDS. Pangenomic analysis clusters them into 12,337 protein clusters (PCs) with a core *Nitrospira* genome consisting of 1,382 PCs. This core genome contains genes involved in the central metabolism of NOB such as the nitrite oxidation pathway and the reductive tricarboxylic acid cycle for CO_2 fixation (rTCA). Other PCs of the core genome are chlorite dismutase, copper-containing nitrite reductase (*nirK*) and the genes of the oxidative TCA cycle. 35 PCs were just found in the comammox *Nitrospira* genomes and 57 and 52 in the comammox clade A and clade B, respectively (**Paper II**). Besides the genes of the *amo* and *hao* operon other genes related to nitrogen were solely found in comammox *Nitrospira*. Thus, comammox genomes contain a broad diversity of urea transporters. Aside from the high affinity urea ABC transporters (*urtABCDE*) present in some *Nitrospira* spp. such as *N. lenta* (Koch *et al.*, 2015), comammox genomes also contain two additional urea transporters: an outer-membrane porin (*fmdC*) involved in uptake of short-chain amides and urea at extremely low concentrations (Mills *et al.*, 1997) and a urea carboxylase-related transporter (*uctT*). Additionally, comammox genomes exclusively harbour an additional copy of agmatinase which hydrolyzes agmatine, producing putrescine and urea. Thus, comammox *Nitrospira* may have a competitive advantage in urea uptake with respect to other *Nitrospira* in environments with low and/or fluctuating urea concentrations. The presence of copper homeostasis genes (*copCD* and *copAB* - with highest sequence similarities to homologs in betaproteobacterial ammonium oxidizers) is another feature distinct for comammox compared to strict nitrite oxidizing *Nitrospira*. These proteins usually confer higher copper tolerance and increased copper uptake (Cha and Cooksey, 1993; Chain *et al.*, 2003), so it could provide an advantage to comammox *Nitrospira* in environments with very high or very low copper availability. In contrast to canonical *Nitrospira*, the comammox genomes harbour a 2/2 hemoglobin type II (TrHb2), which has been associated with both oxidative stress resistance and oxygen scavenging (Ouellet *et al.*, 2007; Torge *et al.*, 2009). On the other hand, opposite to strict nitrite-oxidizing *Nitrospira*, comammox lack some of the genes required for assimilatory nitrite reduction, therefore potentially impeding them to grow with nitrite as the only nitrogen source as was observed for *Ca*. N. inopinata (Daims *et al.*, 2015). Similarly, contrary to canonical *Nitrospira*, cyanate hydratases genes (*cynS*) were not detected in the examined comammox genomes (**Paper II**).

An important difference between clade A and clade B comammox genomes relates to ammonium uptake transporters. While clade A genomes encode a low-affinity Rh-type ammonium transporter homologous to the ones found in betaproteobacterial ammonium oxidizers, clade B genomes harbour the high-affinity AmtB-type transporter. Hence, ammonium uptake affinity disparity might be an important niche-separation factor between the two comammox clades.

4.3 Comparison between comammox and AOP

Ammonia monooxygenase (AMO) and hydroxylamine dehydrogenase, also known to as hydroxylamine oxidoreductase (HAO), are the key enzymes in AOB. The AMO gene cluster consists of *amoCABED* while the HAO cluster is typically formed by *haoABcycAB*. Two or three copies of these clusters are present in the genomes of β-AOB while one copy is present in γ-AOB genomes. AOA encode one copy of AMO but no homologue of HAO has been detected in any AOA genome although hydroxylamine has been shown as an intermediate in their ammonia oxidation pathway (Vajrala *et al.*, 2013). As for comammox *Nitrospira*, one copy of the AMO and HAO clusters were detected in the genomes recovered from the RSF metagenome (**Paper II**) and by others (Daims *et al.*, 2015; van Kessel *et al.*, 2015; Pinto *et al.*, 2015). The existence of multiple copies of these essential enzymes in AOB is poorly understood, but studies in *Nitrosomonas europaea* resulted in different phenotypes when *amoA* and *amoB* were knocked-out (Stein *et al.*, 2000), while it was not the case for *haoA* (Hommes *et al.*, 2002).

Besides differences in AMO and HAO copy numbers, a remarkable dissimilarity was identified in proteins related to NO_x detoxification between comammox and AOB. The membrane-bound cytochrome c nitric oxide reductase (cNOR), the heme-copper nitric oxide reductase (sNOR), the nitrosocyanin, and the cytochrome P460 (*cytL* gene), which are all generally present in AOB, were not detected in any of the comammox *Nitrospira* genomes. This feature is shared with AOA, and might point to an adaptation to environments with low ammonium concentrations. Comammox *Nitrospira* and AOP also differ in the pathway used for CO_2 fixation. While comammox *Nitrospira* contain the microaerophilic reverse tricarboxylic acid cycle (rTCA) pathway, AOB and AOA possess the oxygen-tolerant Calvin–Benson–Bassham and hydroxypropionate–hydroxybutyrate cycles, respectively. Moreover, the canonical AOP genomes commonly encode the low-affinity *aa3*-type haeme-copper oxidase. In contrast, comammox *Nitrospira* harbour high affinity cy-

tochrome *bd*-like oxidases. Furthermore, the 2/2 hemoglobin type II (TrHb2), associated with oxidative stress resistance and oxygen scavenging, which was detected in comammox *Nitrospira* is not universal for AOB (just present in *Nitrosospira* spp) and has not been detected in AOA (Paper II). These three features together could indicate that comammox *Nitrospira* may outcompete AOB and AOA in hypoxic environments.

Comparisons between comammox and AOP genomes also revealed differences related to the capacity to compete under phosphorous and copper limiting conditions. Comammox *Nitrospira* genomes contain an alkaline phosphatase (*phoD*), which has been associated with overexpression under phosphate limitation and starvation (Kageyama *et al.*, 2011; Shen *et al.*, 2016). This enzyme was not detected in AOA genomes and is not universal in AOB (putative homologs are present in *Nitrosomonas sp*. Is79A3*, Nitrosococcus halophilus* and *Nitrosospira* spp.). In relation to Cu^{2+} homeostasis, while *copCD* genes are present in both AOB and comammox genomes, the *copAB* genes detected in comammox *Nitrospira* are not common for AOB as just few species seem to harbour homologues (*Nitrosomonas europaea, Nitrosomonas eutropha, Nitrosospira multiformis and Nitrosospira* sp. Nv17) (Paper II). On the other hand, genes related to Cu^{2+} were not detected in AOA in spite of the high Cu^{2+} demand that these organism required due to their high presence of copper-containing proteins (Amin *et al.*, 2013).

4.4 Evolutionary history of comammox

The presence of ammonium oxidation-related genes in microbes formerly known as strict nitrite oxidizers, raises question whether these genes are ancestral to *Nitrospira* spp. or were acquired by horizontal gene transfer. The degree of divergence of the AMO and HAO sequences with the most closely related AO, β-AOB, makes a recent acquisition improbable, but an earlier gene transfer cannot be ruled out. In fact, we observed that the similarity of AMO and HAO sequence between β-AOB and comammox *Nitrospira* (ca. 60% AmoA and ca. 66% for HaoA) is higher than between β-AOB and γ-AOB (ca. 45% for AmoA and ca. 53% for HaoA), even though these two AOB groups belong to the same phylum (Proteobacteria), and are distantly related to the Nitrospirae phylum, to which *Nitrospira* belong.

In **Paper II**, protein dissimilarity, gene arrangement and reconciliation analyses were performed to unravel the evolutionary history of comammox *Nitrospira*. For the protein dissimilarity analysis, sequences of proteins central to

the ammonium oxidation pathway and 18 housekeeping proteins were compared to that of a set of 14 ribosomal proteins (**Paper II**, Supplementary Table 1). We included sequences from the four recovered genomes in the RSF metagenome together with seven publicly available comammox genomes as well as from eight previously published β-AOB genomes (**Paper II**, Supplementary Table 3). The pairwise dissimilarity comparison showed that while dissimilarity in housekeeping proteins are linearly related to the ribosomal proteins, this is not the case for most of the AO-related proteins. Besides clearly discontinuities, we also detected that the dissimilarities in the AO-related sequences between β-AOB and comammox genomes are similar to dissimilarities between *Nitrospira* comammox clade A and clade B, which makes a recent transfer from β-AOB to a common ancestor of the two clades improbable (Figure 4.3).

Gene agreement analysis revealed common features shared by comammox genomes distinct from canonical AOB: the AMO and HAO genes cluster together location in the shame genomic region with the cytochrome *c* biogenesis genes, as well as a second copy of the *amoD* gene (Paper II for further details). Additionally, the presence of a duplication on one of cytochrome *c* biogenesis genes (*ccml*) in clade B and the existence of a *dctA* and two dupli-

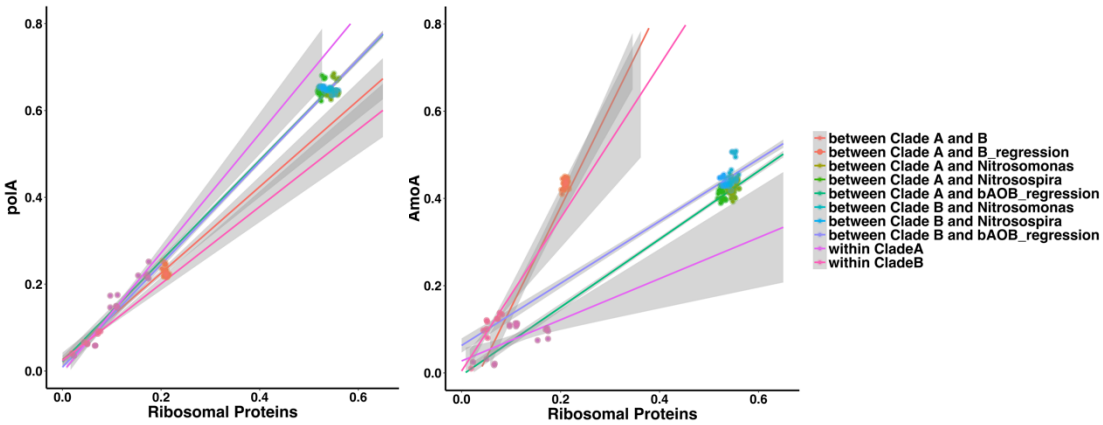

Figure 4.3. Relationship between genomes phylogenetic distance and protein sequence divergence for a house keeping protein (left) and for the ammonium monooxygenase (right) for comammox Nitrospira and β-AOB genomes. Each dot represents a pair of genomes and is coloured according to the groups to which the compared genomes belong. The y-axis shows the pairwise protein dissimilarity (fraction of amino-acid sites that differ) while the x-axis shows the corresponding pairwise dissimilarity for a set of 14 ribosomal proteins. Coloured lines show the linear regression for each groupwise comparison with shadowed regions indicating 95% confidence intervals for the slopes. A: comammox clade A; B: comammox clade B; Ntsm: *Nitrosomonas* genomes; Ntrssp: *Nitrosospira* genomes. AOB: Both *Nitrosomonas* and *Nitrosospira*. **(Paper II)**

cated genes associated with iron storage (*brf*) in the flanking region of the AMO, HAO, cyt. *c* biogenesis clusters found in clade A, distinguish the two comammox clades (Paper II for further details).

For the reconciliation analysis, the 14 ribosomal proteins from the already mention genomes, two γ-AOB and another *Nitrospira* species (Paper II, Supplementary Table 3) were used to construct a species-tree while each AO-related protein sequence were utilized for the gene-tree construction. The probabilistic (Bayesian) analysis provided strong support for horizontal transfer of the majority of the investigated genes into *Nitrospira* (Paper II, Supplementary Table 4). Specifically, there were 12 genes for which the posterior probability for at least one transfer event was 95%-100%, and another 10 for which the posterior probability was 80%-95%. The evolutionary history of *amoA* and *haoA* revealed for *amoA*, two separate early gene transfers from β-AOB to an ancestor of comammox *Nitrospira*, followed by clade specific loss. As a result, the *amoA* genes in clade A and B originate from two separate HGT events (see Paper II for further details). An early transfer of *haoA* from β-AOB to an ancestor of comammox *Nitrospira* is inferred. However, in this case, the higher similarity between comammox clades compared to the similarity of each clade with the β-AOB genomes, suggests one unique horizontal transfer event (see **Paper II** for further details).

Taken together, these analyses suggest horizontal transfer from β-proteobacterial ammonia oxidizers to comammox *Nitrospira* for genes belonging to the ammonium oxidation pathway and consequent loss in *Nitrospira* spp. not containing the AO-related genes.

5 Distribution and abundance of comammox

The genus *Nitrospira* comprises at least six sublineages and is widely distributed in very different environments. *Nitrospira* spp. have been found in soil (Daebeler *et al.*, 2014), freshwater (Hovanec *et al.*, 1998), groundwater (Schwab *et al.*, 2016), seawater (Haaijer *et al.*, 2013) as well as in technical systems as WWTP (Wang *et al.*, 2012) and drinking water systems (Gülay *et al.*, 2016). Additionally, members of several *Nitrospira* lineages were detected in geothermal springs (Edwards *et al.*, 2013). This adaptation to thermophilic environments makes *Nitrospira* the only known NOB with this capacity (Daims *et al.*, 2016).

The finding of microorganisms capable of performing the complete oxidation of ammonium raise the interrogation of how widespread and important this process is in the global nitrogen cycle. Comammox *Nitrospira* appeared as a small fraction of the total community in the mentioned recent discoveries (Chapter 3) – with the notable exception being its high abundance in the rapid sand filter investigated in this thesis (**Paper I**). Whether this apparently low prevalence of comammox is the general trend or whether comammox *Nitrospira* can be a more abundant member - and by deduction, play a more significant role in different environments is still unknown. To address this question, tools to detect and quantify comammox *Nitrospira* in different environments are needed. Out of the six known *Nitrospira* lineages, so far, all described comammox belongs to lineage II (**Paper II**). This lineage, found in a wide variety of different environments, also contains species which are strict nitrite-oxidizers (Daims *et al.*, 2016). In addition, comammox *Nitrospira* do not constitute a monophyletic group within *Nitrospira* lineage II based on the16S rRNA or *nxrB* (nitrite oxidoreductase subunit B) genes phylogenies. Thus, these two genes, which typically used to target *Nitrospira* spp., are not specific enough to exclusively captured comammox *Nitrospira*. In contrast, the AMO identified in both clades of the newly comammox genomes are dissimilar from the other ammonium oxidizers and are not covered by the existing primers targeting the *amoA* genes of AOB or AOA. Hence, AMO would be an ideal target to solely detect comammox *Nitrospira*. In **Paper III**, one of the objective is to develop primers capable of capturing and quantifying comammox *Nitrospira* based on the *amoA* (ammonia monooxygenase subunit A) gene. To do so, DNA sequences for all known, high quality comammox *Nitrospira amoA* genes were collected and used for the design of the primers

which were tested, first in silico and then with biological samples, for coverage and specificity (further details in **Paper III**).

With the new developed primers, we quantified and characterized comammox *Nitrospira* spp. in rapid sand filters from 12 different waterworks using qPCR and amplicon sequencing (Chapter 2). In addition, the same approach was followed to study the whole community as well as specifically other nitrifiers (Table 1 in **Paper III** for the specific primers). The microbial community composition of the filters was consistently dominated by Proteobacteria and Nitrospirae phyla (ranging from 62 and 94%) (Figure 1b in **Paper III**). This microbial structure is similar to those of previously studied groundwater-fed RSFs (Albers *et al.*, 2015; Gülay *et al.*, 2016). *Nitrospira* spp. made up an average of 35% of the total communities based on both qPCR and the sequencing data (Figure 5.1). Based on the cell numbers of total *Nitrospira* (with 16S rRNA and *nxrB* genes) and comammox *Nitrospira* (*Nitrospira amoA*) determined by qPCR, comammox makes up between 40-100% of *Nitrospira* spp. (Figures 5.1, and Figure 2a in **Paper III**). In relation to the two comammox clades, in all the filters clade B was more abundant than clade A (Figure 5.2).

Nitrospira spp. abundance was around one order of magnitude higher than canonical AOB, which comprised less than 3% of the communities, although they reached percentages higher than 10% in few of the filters. AOA were

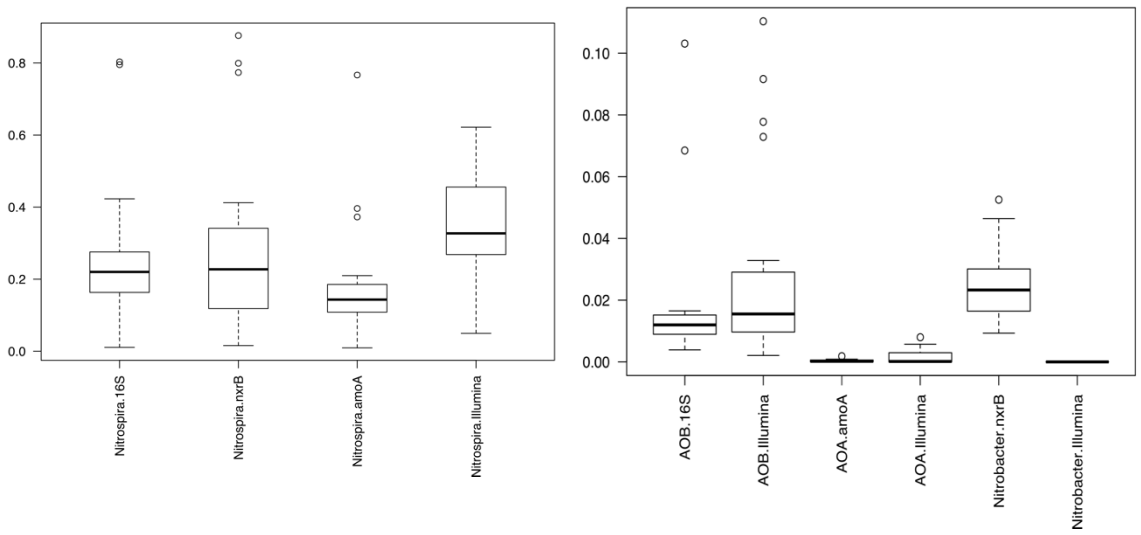

Figure 5.1. Relative abundance of *Nitrospira* spp., AOB, AOA and *Nitrobacter* spp. based on qPCT targeting 16S rRNA and functional genes (*amoA* and *nxrB*), and in 16S rRNA amplicon sequencing. **(Paper III)**

detected in most of the filters even though their abundance was lower than 1% (Figure 5.1).

Taken together, these results indicate that comammox *Nitrospira* are the most abundant nitrifiers in all the examined rapid sand filters treating groundwater.

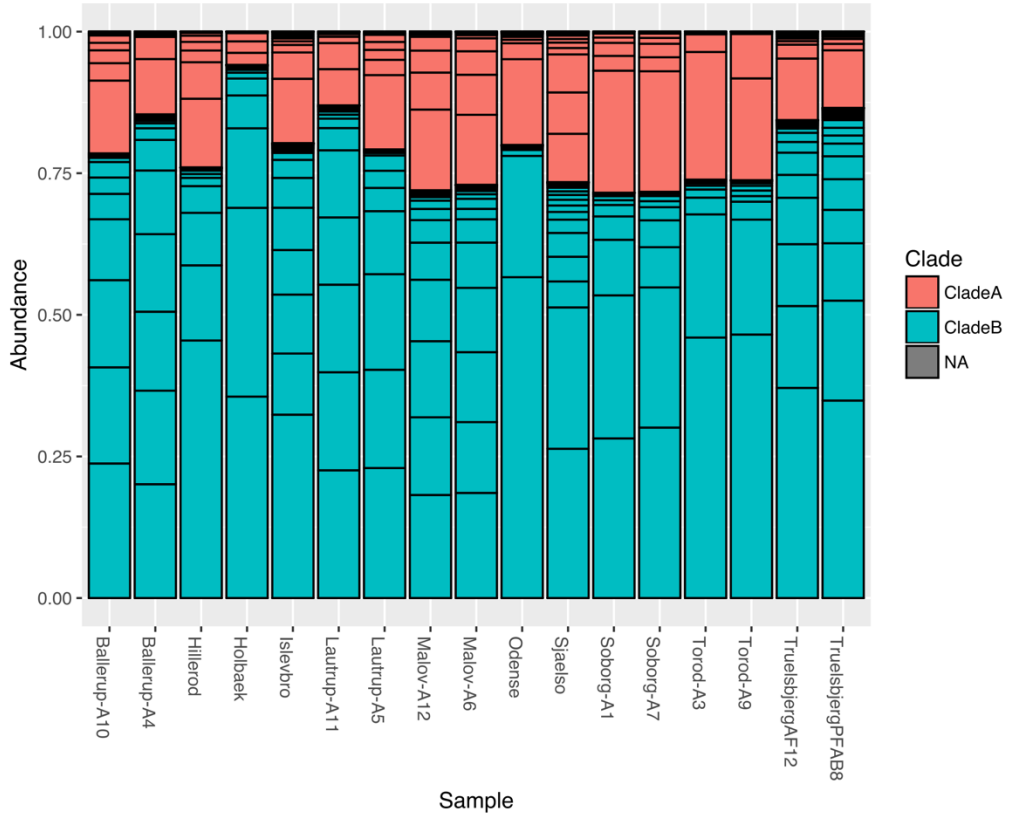

Figure 5.2. Relative abundance of comammox clade A and comammox clade B based on qPCR targeting *amoA* in eleven rapid sand filters. **(Paper III)**

Since the discovery of comammox *Nitrospira*, other research groups have also put efforts in designed primers for the detection of comammox. Thus, Wang *et al*., used a two-step method with primers with broad coverage of copper-containing membrane-bound monoxygenases (CuMMOs) combined with a newly designed specific primer for comammox (Wang *et al.*, 2016). Even though this approach shows a broad coverage for CuMMOs, its coverage of comammox *amoA* is not optimal. Pjevac *et al*., developed set of unique primers for clade A and clade B. Although those primers have a high coverage and specificity simultaneous detection of clades is not possible. Though these primers were subsequently mixed for qPCR, the amplification efficiency may vary, potentially resulting in incorrect ratios of the two clades in quantitative analysis (Pjevac *et al.*, 2016). Another group combined an exist-

ent *pmoA* (particulate methane monooxygenase subunit A) primer with a newly designed comammox specific primer (Bartelme *et al.*, 2017). This approach showed a single comammox sequence amplification. In the absence of further investigation in the coverage of the primer, it remains unclear if this results is due to low coverage of the primer or low diversity of comammox in the investigated samples (Bartelme *et al.*, 2017). In any case, with the potentially future findings of new comammox genomes and *amoA* sequences, refinement of these primers might be needed.

So far, the use of comammox primers together with metagenomics analysis have detected comammox *Nitrospira* in a high variety of different habitat both in engineering and natural systems with exception of marine environments. Specifically, together with the initial discovery of comammox *Nitrospira* in the pipe of an oil exploration (Daims *et al.*, 2015), a trickling filter connected to a recirculation aquaculture system (van Kessel *et al.*, 2015), a drinking water distribution (Pinto *et al.*, 2015), and groundwater fed rapid sand filter (Paper I); comammox have been also identified in wastewater treatment plants (Chao *et al.*, 2016), fine particulate matter (Gao *et al.*, 2016), rice paddy and forest soils, brackish lake sediment, and in a freshwater biofilm (Pjevac *et al.*, 2016).

6 Ecophysiology of comammox

Prokaryotes within a functional group have been frequently detected co-occurring in the same environment (Tessier and Leibold, 1997; Haverkamp *et al.*, 2009; Gonzaga *et al.*, 2012). When this co-occurrence goes beyond the merely temporal occupancy of the same habitat, then niche differentiation is needed. This coexistence of functionally similar species is often possible because of divergent physiological requirements and interaction with the surrounding environment (Leibold and McPeek, 2006). Since the discovery of AOA, an inquiry about their role in ammonia oxidation in comparison with AOB started. The extensive detection of both nitrifying groups in the same environment (Yamamoto *et al.*, 2010; Wessén *et al.*, 2010; Vissers *et al.*, 2013) led to investigate which distinguishing characteristics could drive their coexistence. Correlation studies and experimental approaches have pointed towards different factors which could be involved in AOB and AOA niche partitioning; the suggested distinctions have been differential optima for ammonia concentrations and pH, among others (Prosser and Nicol, 2012). Now, with the discovery of comammox *Nitrospira*, another group of ammonia oxidizers muddles the picture. No studies have investigated the reasons for occurrence of comammox in the environment but the simultaneous presence of comammox, AOB and AOA in the same habitat has already been reported (Palomo *et al.*, 2016; Bartelme *et al.*, 2017; Wang *et al.*, 2017) (**Paper I**).

Several methods have emerged that allow simultaneous detection of in situ activity and taxonomic identification at either the community-level or at single-cell resolution. Most of them rely on the principle of exposing microbial communities to substrates labeled with stable or radioactive isotopes. The active microorganisms which utilize the labeled substrate during growth will incorporate the isotopes into their biomolecules. Two of these techniques are stable-isotope-probing (SIP) and microautoradiography (MAR). SIP can be combined with extraction and separation of labeled biomolecules as DNA and RNA for the taxonomic identification of the active community. This approach was used to investigate which organic substrates were able to utilize sulfate-reducers in marine sediments. It was detected that *Desulfobacteraceae* spp. were important consumers of propionate but not of glucose in that habitat (Miyatake *et al.*, 2009). In the case of microautoradiography, the taxonomic identification is achieved by combining the addition of a radiolabeled substrate with fluorescent in situ hybridization (MAR-FISH). *Nitrospira*-like bacteria were confirmed to be able to grow mixotrophically in the presence of

pyruvate by using MAR-FISH (Daims *et al.*, 2001). More recent techniques to link microbial activity with specific taxonomic groups are NanoSIMS and Raman microspectroscopy (Wagner, 2009). NanoSIMS, which is based on an advanced type of dynamic secondary ion mass spectrometer, has been successfully coupled with in situ hybridization to detect that the anaerobic and phototropic species *Chromatium okenii*, accounting for less than 1% of the total community abundance, could contribute to more than 40% of the total ammonium and carbon uptake in an oligotrophic lake (Musat *et al.*, 2008). Raman microspectroscopy enables the determination of chemical bonding patterns related to biological molecules within individual cells. The combination of Raman with FISH showed that an uncultured *Acidovorax* sp., played an essential role in naphthalene biodegradation in groundwater (Huang *et al.*, 2009).

These techniques could have some limitations in samples including comammox *Nitrospira* with other nitrifiers. In the case of SIP, label N has some difficulties because most of the end products are not assimilated (Pratscher *et al.*, 2011), Additionally, inhibitors as allylthiourea (ATU) and chlorate can encounter analytical interferences and their selectiveness for comammox *Nitrospira* is still unknown, although first experiments suggest challenges for samples containing comammox as well as canonical NOB and AOB (Tatari *et al.*, 2017). In the case of FISH-based methods, specific 16S rRNA-based probes for comammox *Nitrospira* could be challenging as these microorganisms are not a monophyletic group.

Here, we used a previously optimized lab-scale assay that can mimic full-scale rapid sand filter conditions at controlled ammonium loadings (Tatari *et al.*, 2013) (Figure 6.1). We investigated the response of comammox *Nitrospira*, AOB and AOA under different conditions including ammonium loading rate, oxygen limitation, temperature and addition of an external carbon source (**Paper IV**). The lab-scale experiment comprised small plexiglass columns (5 cm bed height, 2.6 cm inner diameter) packed with sand and continuously fed with filter influent or effluent water. The used filter was collected from a RSF at Islevbro waterworks which contains different relative proportions of comammox and canonical AOP and NOB (**Paper I**).

Two set of experiments with eight parallel columns were run for 30 days (Table 6.1) at 20°C and at 10°C, respectively. In the experiment 1, columns were filled with top or bottom filter material and fed with effluent water from the investigated sand filter spiked with ammonium. The reference loading condi-

tion was 35 g NH_4^+-N/m^3 filter material/d (1 mg/L NH_4^+-N at an influent flowrate of 0.96 L/d) mimicking full scale conditions. In experiment 1, besides the reference loading rate (1 Loading Rate; 1*L_{NH4+}), columns were exposed to 0.1*L_{NH4+} and 5*L_{NH4+} (with and without oxygen limitation). In experiment 2, the influence of an additional carbon and energy source (effluent water spiked with 1mg/L of acetate or influent water from the RSF; with and without oxygen limitation) in the microbial communities was evaluated. In all the columns, the volumetric ammonium loading rate increased gradually during the experiment as 2 g of filter material were removed from the top of each column (four times in the 30 days).

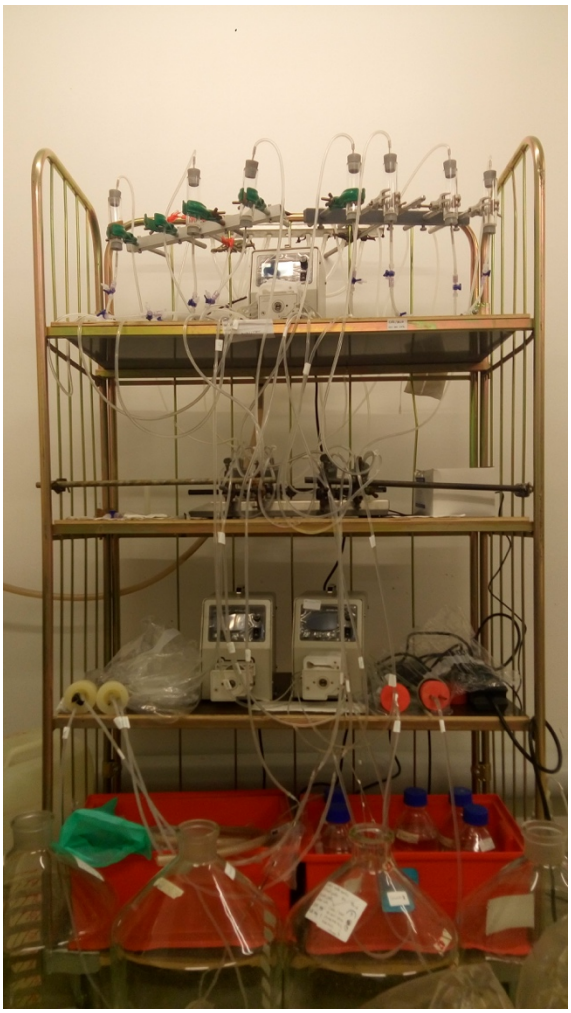

Figure 6.1. Lab-scale set up mimicking a full-scale rapid sand filter. Top part: eight columns. Middle-top part: recirculation system: Middle-bottom part: pumps; Bottom part: influent bottles.

Table 6.1. Summary of experimental design.

Experiment	Condition	T (ºC)	Origin (cm)	Loading (NH_4^+)	Medium	Oxygen Limitation	ID	Removal efficiency
1	1	20	0-10	0.1*L	E	NO	0.1*L E 20C T	87 ± 4
	2		40-50	0.1*L		NO	0.1*L E 20C B	77 ± 20
	3		0-10	5*L		Yes	5*L E 20C T O2	24 ± 17
	4		40-50	5*L		Yes	5*L E 20C B O2	11 ± 4
	5		0-10	1*L		NO	1*L E 20C T	93 ± 7
	6		40-50	1*L		NO	1*L E 20C B	44 ± 13
	7		0-10	5*L		NO	5*L E 20C T	93 ± 2
	8		40-50	5*L		NO	5*L E 20C B	34 ± 2
2	9	10	0-10	1*L	E	NO	1*L E 10C	98 ± 2
	10 (n =2)			1*L	I	NO	1*L I 10C	99 ± 1
	11 (n =2)			1*L	E + A	NO	1*L E+A 10C	98 ± 3
	12 (n =2)			5*L	I	Yes	5*L I 10C O2	44 ± 10
	13			5*L	E + A	Yes	5*L E+A 10C O2	43 ± 7

In the top layer columns, nearly complete ammonium removal was observed at all loading conditions except under oxygen limitation. In contrast, incomplete ammonium removal was detected for all the bottom layer columns (Table 1 and **Paper IV**, Figure 1). Removal efficiency increased after the first days in the bottom layer columns. Nitrite accumulation was only detected in the columns running at oxygen limitation during the first hours of the experiment.

The main difference in the microbial communities between the initial top and bottom layer material was the relative abundance of *Nitrospira* spp, accounting for 28.8±1.2% and 3.9% of total community the top and bottom layer, respectively (**Paper IV**, Figure 2). In turn, *Nitrosomonas* spp. densities were at 0.9±0.6% and 0.1% while AOA were a 0.03±0.01% and 0.1% in the top and bottom layer, respectively.

Microbial growth was especially observed in the bottom layer columns with an average increase of 34% in the total cell density. AOB and *Nitrospira* spp. moderately increased in all the columns of the experiment 2 and especially in the bottom layer columns of the experiment 1. Additionally, AOB and *Nitrospira* spp. showed a higher fitness than the rest of the community particularly in the columns fed with effluent water and $1*L_{NH4+}$ or under oxygen limita-

tion (**Paper IV**). AOA did not significantly vary in absolute cell density in experiment 1 but had a large increase in all the columns in the experiment at 10°C.

The 16S rRNA amplicon libraries revealed an increment in *Nitrospira* spp. in the bottom layer columns. Moreover, we detected a general increase in a comammox clade A sequence (100% identical to the populated genome CG24B, **Paper II**) at all conditions in the experiments run at 10°C, together with a simultaneous decrease in a *Nitrospira* sp. sequence which was most abundant the beginning of the experiment (100% identical to the population genome CG24A, **Paper II**). A possible explanation for this observation is a better adaptation of the comammox clade A sequence (CG24B) to increasing ammonium loading as even in the columns mimicking the full-scale RSF conditions the volumetric loading rate increased four times during the experiment. Interestingly, in contrast to clade A comammox genomes, clade B genomes (CG24A) harbour high affinity ammonium transporters (**Paper II**) which may be advantageous at lower ammonium conditions.

The increased proportion of comammox *Nitrospira* sequences within the total *Nitrospira* sequences after the 30 days experiment together with the observation that total *Nitrospira* cell density increment was always much higher than the sum of the AOB an AOA cell number increase, strongly suggest that *Nitrospira* is involved in ammonium oxidation under the tested conditions.

Besides these targeted lab-scale experiments, we also used the comparative genomic analysis to get insights into the ecophysiology of comammox *Nitrospira*. This approach has been previously used to illuminate common and differentiating features in closely related or co-occurring microorganisms. Kato *et al.* (2015) detected that different neutrophilic iron oxidizers contain distinctions in carbon, nitrogen and sulfur cycling genes which could separate them in particular niches (Kato *et al.*, 2015). Similarly, a comparative genomics analysis on four haloarchaea from a cold and hypersaline lake in Antarctica revealed distinct preferences for carbon substrates which could allow their co-occurrence throughout the water column (Williams *et al.*, 2014).

The comparative genomics of 16 *Nitrospira* genomes highlighted genomic characteristics (described in Chapter 4 and further discussed in **Paper II**) which could lead to niche differentiation between the *Nitrospira* types. Thus, the greater variety of urea transporter in comammox *Nitrospira* hints at a possible competitive advantage in urea uptake with respect to other *Nitrospira* in environments with low and/or fluctuating urea concentrations. In addi-

tion, the high presence of copper homeostasis genes could provide an advantage to comammox *Nitrospira* in environments with very high or very low copper availability. The distinct ammonium uptake affinity systems in comammox clade A and clade B could be essential in their niche-separation.

The decreased presence of NO_x detoxification pathways in comammox *Nitrospira* and AOA suggests a better adaptation for environments with low ammonium concentrations in comparison with canonical AOB. Moreover, comammox *Nitrospira* could outcompete AOB and AOA under phosphorous and copper limiting conditions thank to their higher number of copper homeostasis genes as well as an alkaline phosphatase (*phoD*) detected to be highly expressed under phosphorus limitation and starvation (Shen *et al.*, 2016).

7 Conclusions

This PhD project has contributed to the discovery of complete ammonium oxidizing microorganisms (comammox). This discovery forces a revision of one of the most important biogeochemical cycles on Earth, the nitrogen cycle. This discovery provides closure to the unexpected previous observations of high abundances of *Nitrospira* spp. in rapid sand filter microbial communities. Additionally, this thesis has provided the first extensive comparative description of the genomic capabilities of comammox *Nitrospira* in relation to canonical ammonium and nitrite oxidizers.

The main findings of this work are summarized below:

- Through a metagenomic analysis and a genome recovery approach, 14 near-complete population genomes were reconstructed from a groundwater-fed rapid sand filter community and functionally annotated. These organisms have the genetic capacity to grow on the typical compounds found in the source water. Hence, we identified population genomes with capabilities to oxidize ammonium, nitrite, methane, hydrogen sulfide, iron and manganese as well as to assimilate organic compounds. Additionally, a highly abundant *Nitrospira* composite population genome contained the genes for complete ammonium oxidation (comammox). Moreover, a canonical ammonium oxidizing bacteria population genome was recovered and genes belonging to ammonium oxidizing archaea were also detected. These observations point towards co-existence of multiple nitrifiers within the same biofilter.
- A differential coverage binning method allowed the separation of the *Nitrospira* composite genome into five individual population genomes. Four of the recovered genomes are comammox *Nitrospira*. The comparative analysis of these, along with high-quality published *Nitrospira* genomes, revealed distinguishing genetic capabilities for the two comammox clades, canonical *Nitrospira* and strict ammonia oxidizers. Divergences were detected for nitrogen source utilization capacity, electron donor versatility and stress response inventory, among others. Specifically, comammox *Nitrospira* harbour a high variety of genes related with tolerance to low substrate availability. In addition, the analysis of the evolutionary history of comammox *Nitrospira* indicates acquisition by horizontal gene transfer of several genes involved in the ammonium oxidation pathway from betaproteobacterial AOB. In some of those genes

we also detected transfer events from comammox clade B to clade A genomes.

- Primers targeting comammox *Nitrospira* were developed for exclusive detection of these microorganisms. Exploration of RSF communities from 12 drinking water treatment plants across Denmark disclosed that comammox *Nitrospira* are the most abundant nitrifiers in the investigated groundwater-fed rapid sand filters. In turn, canonical AOB and AOA accounted for an average of 3% and less than 1%, respectively. In all the examined biofilters, comammox clade B was dominant over clade A.
- The effect of different ammonium loading conditions as well as the presence of an external carbon source and oxygen limitation on filter material with different relative proportions of nitrifiers was investigated using a lab-scale experimentation under conditions that mimic full-scale RSF. Simultaneous growth of comammox *Nitrospira*, AOB and AOA was detected. No immediate insights in conditions selecting for comammox over the other nitrifiers were identified. However, increasing loading rates in the microbial communities from top layer material selected comammox clade A over clade B. Moreover, AOA had a lower fitness at high temperature in comparison with the other nitrifiers. Lastly, significant growth of comammox *Nitrospira* and AOB was measured in the experiment initiated with bottom layer material.

8 Future perspectives

In this PhD thesis 14 near-complete population genomes were successfully recovered from a rapid sand filter. Their functional annotation allows to predict a model with the metabolic and geochemical processes likely occurring in the filter and the putative involvement of each population genome in the degradation of the groundwater components. However, the genomic analysis was based on DNA which just shows the potential of the community. Studies at expression (metatransciptomics; initial experiments in **Appendix I**) or protein (metaproteomics) level could better relate the specific genomes with the in situ activity. Additionally, these approaches together with other techniques linking presence with activity could unravel if the high *Nitrospira* abundance is solely due to the nitrite and putative ammonium oxidation or if genes and proteins belonging to other pathways are expressed or translated as well.

Although five *Nitrospira* genomes could be successfully recovered from the metagenome, the co-occurrence of closely similar species makes it difficult to completely rule out the existence of heterogeneity in the population genomes. Efforts in enrichments or combination of fluorescence-activated cell sorting with single-cell sequencing could counteract the potential micro-diversity issue (initial experiments in **Appendix II**). Enrichments are not only helpful for purest genome recovery but also to study physiology of comammox *Nitrospira*. The lab-scale assay showed high potential for enrichments as well as competition studies, however several improvements can be done in future experiments. The higher changes in the nitrifying and total microbial community observed in the bottom layer columns suggests higher potential for competition experiments. In contrast, the initial high abundance of specific comammox sequences in the top later material indicates that easier enrichment could be obtained with this material. Still, conditions for specific selection of clades and particular strains over others are required. In general, longer times could increase the trends and sharper deviation in the loading conditions could enhance niche discrimination. Moreover, the comparative genomics analysis could be a good starting point to test exclusive potential functions distinctly detected in singular genomes. For comammox *Nitrospira* enrichments purpose, besides initial use of ampicillin reported as *Nitrospira*-selective over *Nitrobacter* and heterotrophs (Spieck *et al.*, 2006), low concentrations of ammonium are proposed for outcompeting AOB. In addition, the genomic capacity of comammox *Nitrospira* to better resist under copper and

phosphorous limitations makes effluent over influent water as preferable source water.

Reconciliation analysis enabled to shed light on comammox *Nitrospira* evolutionary history. Nevertheless, the probable future discovery of more and distant comammox genomes will facilitate an even better analysis with reduced uncertainty, might allow to assign actual time to the transfer events and try to reconstruct the whole evolutionary history for all the ammonium oxidation-related genes. The design and development of comammox-specific primers allowed the detection of comammox *Nitrospira* in multiple rapid sand filter and showed the high abundance of comammox in all the biofilters. For future experiments, selection of waterworks with more diverse groundwater constituents may reveal higher variation in the nitrifiers relative proportions and support multivariate analysis to correlate comammox, AOB or AOA abundances with specific physicochemical parameters.

The discovery of comammox *Nitrospira* has, once again, modified our insight of the nitrogen cycle. Many N conversion processes observed in natural and engineered systems may have to be re-evaluated to assess the potential role and interference of the comammox metabolism. For example, in nitritation/anammmox-based technologies for N removal from wastewaters, efforts are usually needed to suppress canonical NOB to enhance process performance (Lotti *et al.*, 2014): indeed, aerobic nitrite to nitrate conversion is undesirable. In that sense, an organism that performs comammox would seem undesirable in such technologies. Yet, comammox bacteria has been observed in close association with anammox bacteria (van Kessel *et al.*, 2015) in enrichment cultures hinting at other, possibly beneficial, interactions. Careful experiments on these systems are needed to elucidate the relation between these microorganisms and the potential advantages in the process performance. N_2O is an important greenhouse gas, and is produced by heterotrophic organisms as well as AOB and AOA (Braker and Conrad, 2011), and as such form a challenge in many N removal technologies. Genes related to N_2O production have been not detected in comammox *Nitrospira* (**Paper II**), providing potential benefits of using comammox based biotechnologies for N removal. Experiments clarifying the N_2O production potential, or lack thereof, by comammox are required, in order to assess their contribution to global N_2O production, and benchmark them against AOB and AOA for technical applications. Clearly, the contribution potential of comammox *Nitrospira* to novel biotechnologies exists, but awaits further investigation.

9 References

Albers CN, Ellegaard-Jensen L, Harder CB, Rosendahl S, Knudsen BE, Ekelund F, *et al.* (2015). Groundwater Chemistry Determines the Prokaryotic Community Structure of Waterworks Sand Filters. *Environ Sci Technol* **49**: 839–846.

Albertsen M, Hansen LBS, Saunders AM, Nielsen PH, Nielsen KL. (2012). A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J* **6**: 1094–106.

Ali M, Okabe S. (2015). Anammox-based technologies for nitrogen removal: Advances in process start-up and remaining issues. *Chemosphere* **141**: 144–153.

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, *et al.* (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.

Amann RI, Ludwig W, Schleifer KH. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–69.

Amin S a., Moffett JW, Martens-Habbena W, Jacquot JE, Han Y, Devol A, *et al.* (2013). Copper requirements of the ammonia-oxidizing archaeon Nitrosopumilus maritimus SCM1 and implications for nitrification in the marine environment. *Limnol Oceanogr* **58**: 2037–2045.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aoi Y, Kinoshita T, Hata T, Ohta H, Obokata H, Tsuneda S. (2009). Hollow-Fiber Membrane Chamber as a Device for In Situ Environmental Cultivation. *Appl Environ Microbiol* **75**: 3826–3833.

Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. (2015). Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* **31**: 2882–2884.

Bartelme RP, McLellan SL, Newton RJ. (2017). Freshwater Recirculating Aquaculture System Operations Drive Biofilter Bacterial Community Shifts Around a Stable Nitrifying Consortium of Ammonia-oxidizing Archaea and Comammox Nitrospira. *Front Microbiol* **8**: 101.

Beman JM, Leilei Shih J, Popp BN. (2013). Nitrite oxidation in the upper water column and oxygen minimum zone of the eastern tropical North Pacific Ocean. *ISME J* **7**: 2192–205.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Braker G, Conrad R. (2011). Diversity, structure, and size of N 2O-producing microbial communities in soils-what matters for their functioning?

Broda E. (1977). Two kinds of lithotrophs missing in nature. *Z Allg Mikrobiol* **17**: 491–493.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, *et al.* (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208–211.

Bryant D a, Costas AMG, Maresca J a, Chew AGM, Klatt CG, Bateson MM, *et al.* (2007).

Candidatus Chloracidobacterium thermophilum: an aerobic phototrophic Acidobacterium. *Science* **317**: 523–6.

Cai Y, Li D, Liang Y, Luo Y, Zeng H, Zhang J. (2015). Effective start-up biofiltration method for Fe, Mn, and ammonia removal and bacterial community analysis. *Bioresour Technol* **176**: 149–155.

Canfield DE, Glazer AN, Falkowski PG. (2010). The Evolution and Future of Earth's Nitrogen Cycle. *Science (80-)* **330**: 192–196.

Carrión O, Curson a. RJ, Kumaresan D, Fu Y, Lang a. S, Mercadé E, *et al.* (2015). A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nat Commun* **6**: DOI: 10.1038/ncomms7579.

Cassman KG, Dobermann A, Walters DT. (2002). Agroecosystems, Nitrogen-use Efficiency, and Nitrogen Management. *AMBIO A J Hum Environ* **31**: 132–140.

Cerrato JM, Falkinham JO, Dietrich AM, Knocke WR, McKinney CW, Pruden A. (2010). Manganese-oxidizing and -reducing microorganisms isolated from biofilms in chlorinated drinking water systems. *Water Res* **44**: 3935–45.

Cha JS, Cooksey DA. (1993). Copper hypersensitivity and uptake in Pseudomonas syringae containing cloned components of the copper resistance operon. *Appl Environ Microbiol* **59**: 1671–1674.

Chain P, Lamerdin J, Larimer F, Regala W, Lao V, Land M, *et al.* (2003). Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea †. *J Bacteriol* **185**: 2759–2773.

Chao Y, Mao Y, Yu K, Zhang T. (2016). Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach. *Appl Microbiol Biotechnol* **100**: 8225–8237.

Chu C, Lu C, Lee C. (2005). Effects of inorganic nutrients on the regrowth of heterotrophic bacteria in drinking water distribution systems. *J Environ Manage* **74**: 255–263.

Clément JC, Shrestha J, Ehrenfeld JG, Jaffé PR. (2005). Ammonium oxidation coupled to dissimilatory reduction of iron under anaerobic conditions in wetland soils. *Soil Biol Biochem* **37**: 2323–2328.

Compeau PEC, Pevzner PA, Tesler G. (2011). How to apply de Bruijn graphs to genome assembly. *Nat Biotechnol* **29**: 987–991.

Costa E, Pérez J, Kreft J-U. (2006). Why is metabolic labour divided in nitrification? *Trends Microbiol* **14**: 213–219.

Crittenden J, Trussell R, Hand D, Howe K, Tchobanoglous G. (2005). Water Treatment Principles and Design. 2nd ed. Hoboken, N.J : J. Wiley.

Daebeler A, Le Bodelier P, Yan Z, Hefting MM, Jia Z, Laanbroek HJ. (2014). Interactions between Thaumarchaea, Nitrospira and methanotrophs modulate autotrophic nitrification in volcanic grassland soil. *ISME J* **8**: 2397–2410.

Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, *et al.* (2015). Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.

Daims H, Lücker S, Wagner M. (2016). A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol* **24**: 699–712.

Daims H, Nielsen JL, Nielsen PH, Schleifer KH, Wagner M. (2001). In Situ Characterization of Nitrospira-Like Nitrite-Oxidizing Bacteria Active in Wastewater Treatment Plants. *Appl Environ Microbiol* **67**: 5273–5284.

Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**: e243.

Dechesne A, Musovic S, Palomo A, Diwan V, Smets BF. (2016). Underestimation of ammonia-oxidizing bacteria abundance by amplification bias in amoA-targeted qPCR. *Microb Biotechnol* **9**: 519–524.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

Duce R a, LaRoche J, Altieri K, Arrigo KR, Baker a R, Capone DG, *et al.* (2008). Impacts of atmospheric anthropogenic nitrogen on the open ocean. *Science* **320**: 893–7.

Edwards TA, Calica NA, Huang DA, Manoharan N, Hou W, Huang L, *et al.* (2013). Cultivation and characterization of thermophilic Nitrospira species from geothermal springs in the US Great Basin, China, and Armenia. *FEMS Microbiol Ecol* **85**: 283–292.

Erisman JW, Galloway JN, Seitzinger S, Bleeker A, Dise NB, Petrescu AMR, *et al.* (2013). Consequences of human modification of the global nitrogen cycle. *Philos Trans R Soc Lond B Biol Sci* **368**: 20130116.

Erisman JW, Sutton M a., Galloway J, Klimont Z, Winiwarter W. (2008). How a century of ammonia synthesis changed the world. *Nat Geosci* **1**: 636–639.

Espejo-Herrera N, Gràcia-Lavedan E, Boldo E, Aragonés N, Pérez-Gómez B, Pollán M, *et al.* (2016). Colorectal cancer risk and nitrate exposure through drinking water and diet. *Int J Cancer* **139**: 334–346.

Ettwig KF, Van Alen T, Van De Pas-Schoonen KT, Jetten MSM, Strous M. (2009). Enrichment and molecular detection of denitrifying methanotrophic bacteria of the NC10 phylum. *Appl Environ Microbiol* **75**: 3656–3662.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MMM, *et al.* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.

Ettwig KF, Shima S, Van De Pas-Schoonen KT, Kahnt J, Medema MH, Op Den Camp HJM, *et al.* (2008). Denitrifying bacteria anaerobically oxidize methane in the absence of Archaea. *Environ Microbiol* **10**: 3164–3173.

Ettwig KF, Speth DR, Reimann J, Wu ML, Jetten MSM, Keltjens JT. (2012). Bacterial oxygen production in the dark. *Front Microbiol* **3**: 1–8.

Feng S, Xie S, Zhang X, Yang Z, Ding W, Liao X, *et al.* (2012). Ammonium removal pathways and microbial community in GAC-sand dual media filter in drinking water treatment. *J Environ Sci (China)* **24**: 1587–1593.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Foerstner KU, von Mering C, Hooper SD, Bork P. (2005). Environments shape the nucleotide composition of genomes. *EMBO Rep* **6**: 1208–1213.

Foesel BU, Gieseke A, Schwermer C, Stief P, Koch L, Cytryn E, *et al.* (2008). Nitrosomonas Nm143-like ammonia oxidizers and Nitrospira marina-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol Ecol* **63**: 192–204.

Francis CA, Beman JM, Kuypers MM. (2007). New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *Isme J* **1**: 19–27.

Füssel J, Lam P, Lavik G, Jensen MM, Holtappels M, Günter M, *et al.* (2012). Nitrite oxidation in the Namibian oxygen minimum zone. *ISME J* **6**: 1200–1209.

Galloway JN, Aber JD, Erisman JW, Seitzinger SP, Howarth RW, Cowling EB, *et al.* (2003). The Nitrogen Cascade. *Am Inst Biol Sci* **53**: 341.

Galloway JN, Cowling EB. (2002). Reactive Nitrogen and The World: 200 Years of Change. *AMBIO A J Hum Environ* **31**: 64–71.

Gao J-F, Fan X-Y, Pan K-L, Li H-Y, Sun L-X. (2016). Diversity, abundance and activity of ammonia-oxidizing microorganisms in fine particulate matter. *Sci Rep* **6**: 38785.

Ghurye JS, Cepeda-Espinoza V, Pop M. (2016). Metagenomic assembly: Overview, challenges and applications. *Yale J Biol Med* **89**: 353–362.

Glusman G, Severson A, Dhankani V, Robinson M, Farrah T, Mauldin DE, *et al.* (2015). Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet* **6**. e-pub ahead of print, doi: 10.3389/fgene.2015.00045.

Gonzaga A, Martin-Cuadrado AB, López-Pérez M, Mizuno CM, García-Heredia I, Kimes NE, *et al.* (2012). Polyclonality of concurrent natural populations of Alteromonas macleodii. *Genome Biol Evol* **4**: 1360–1374.

Van de Graaf AA, Mulder A, De Bruijn P, Jetten MSM, Robertson LA, Kuenen JG. (1995). Anaerobic oxidation of ammoinium is a biologically mediated process. *Appl Environ Microbiol* **61**: 1246–1251.

Griffin BM, Schott J, Schink B. (2007). Nitrite, an electron donor for anoxygenic photosynthesis. *Science* **316**: 1870.

Gruber-Dorninger C, Pester M, Kitzinger K, Savio DF, Loy A, Rattei T, *et al.* (2014). Functionally relevant diversity of closely related Nitrospira in activated sludge. *ISME J* **9**: 643–55.

Gülay A, Musovic S, Albrechtsen H-J, Al-Soud WA, Sørensen SJ, Smets BF. (2016). Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. *ISME J* **10**: 2209–2222.

Gydesen E, Tkker H. (2013). Groundwater Chemistry and Treatment: Application to Danish Waterworks. In: *Water Treatment*. InTech. e-pub ahead of print, doi: 10.5772/54166.

Haaijer SCM, Ji K, Niftrik L van, Hoischen A, Speth D, Jetten MSM, *et al.* (2013). A novel marine nitrite-oxidizing Nitrospira species from Dutch coastal North Sea water. *Front Microbiol* **4**. e-pub ahead of print, doi: 10.3389/fmicb.2013.00060.

Haft DH. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**: 371–373.

Haverkamp TH a, Schouten D, Doeleman M, Wollenzien U, Huisman J, Stal LJ. (2009). Colorful microdiversity of Synechococcus strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* **3**: 397–408.

Hofman T, Lees H. (1953). The biochemistry of the nitrifying organisms. III. Composition of Nitrosomonas. *Biochem J* **54**: 293–295.

Hommes NG, Sayavedra-Soto LA, Arp DJ. (2002). The roles of the three gene copies encoding hydroxylamine oxidoreductase in Nitrosomonas europaea. *Arch Microbiol* **178**: 471–476.

Hovanec TA, Taylor LT, Blakis A, Delong EF. (1998). Nitrospira-like bacteria associated with nitrite oxidation in freshwater aquaria. *Appl Environ Microbiol* **64**: 258–64.

Huang S, Chen C, Peng X, Jaffé PR. (2016). Environmental factors affecting the presence of Acidimicrobiaceae and ammonium removal under iron-reducing conditions in soil environments. *Soil Biol Biochem* **98**: 148–158.

Huang S, Jaffé PR. (2015). Characterization of incubation experiments and development of an enrichment culture capable of ammonium oxidation under iron-reducing conditions. *Biogeosciences* **12**: 769–779.

Huang WE, Ferguson A, Singer AC, Lawson K, Thompson IP, Kalin RM, *et al.* (2009). Resolving genetic functions within microbial populations: In situ analyses using rRNA and mRNA stable isotope probing coupled with single-cell raman-fluorescence in situ hybridizationδ. *Appl Environ Microbiol* **75**: 234–241.

Huson DH, Weber N. (2013). Microbial Community Analysis Using MEGAN. 1st ed. Elsevier Inc.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, *et al.* (2016). Piphillin: Improved prediction of metagenomic content by direct inference from human microbiomes. *PLoS One* **11**: 1–18.

Jones RR, Weyer PJ, Dellavalle CT, Inoue-Choi M, Anderson KE, Cantor KP, *et al.* (2016). Nitrate from drinking water and diet and bladder cancer among postmenopausal women in Iowa. *Environ Health Perspect* **124**: 1751–1758.

Juretschko S, Timmermann G, Schmid M, Schleifer KH, Pommerening-Röser A, Koops HP, *et al.* (1998). Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. *Appl Environ Microbiol* **64**: 3042–51.

Kaeberlein T. (2002). Isolating 'Uncultivable' Microorganisms in Pure Culture in a Simulated Natural Environment. *Science (80-)* **296**: 1127–1129.

Kageyama H, Tripathi K, Rai AK, Cha-Um S, Waditee-Sirisattha R, Takabe T. (2011). An alkaline phosphatase/phosphodiesterase, PhoD, induced by salt stress and secreted out of the cells of Aphanothece halophytica, a halotolerant cyanobacterium. *Appl Environ Microbiol* **77**: 5178–5183.

Kang DD, Froula J, Egan R, Wang Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.

Karlin S, Mrázek J, Campbell AM. (1997). Compositional biases of bacterial genomes and evolutionary implications. *J Bacteriol* **179**: 3899–3913.

Karst SM, Kirkegaard RH, Albertsen M. (2016). Mmgenome: a Toolbox for Reproducible Genome Extraction From Metagenomes. *bioRxiv* 59121.

Kartal B, Kuenen JG, van Loosdrecht MCM. (2010). Sewage Treatment with Anammox. *Science (80-)* **328**: 702–703.

Kartal B, Maalcke WJ, de Almeida NM, Cirpus I, Gloerich J, Geerts W, *et al.* (2011). Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**: 127-U159.

Kato S, Ohkuma M, Powell DH, Krepski ST, Oshima K, Hattori M, *et al.* (2015). Comparative genomic insights into ecophysiology of neutrophilic, microaerophilic iron oxidizing bacteria. *Front Microbiol* **6**: 1–16.

van Kessel MAHJ, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, *et al.* (2015). Complete nitrification by a single microorganism. *Nature* **528**: 555–559.

Kim M, Lee K-H, Yoon S-W, Kim B-S, Chun J, Yi H. (2013). Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics Inform* **11**: 102–13.

Kinnunen M, Gülay A, Albrechtsen H-J, Dechesne A, Smets BF. (2017). Nitrotoga is selected over Nitrospira in newly assembled biofilm communities from a tap water source community at increased nitrite loading. *Environ Microbiol* **In press**. e-pub ahead of print, doi: 10.1111/1462-2920.13792.

Koch H, Galushko A, Albertsen M, Schintlmeister A, Spieck E, Richter A, *et al.* (2014). Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**: 1052–4.

Koch H, Lücker S, Albertsen M, Kitzinger K, Herbold C, Spieck E, *et al.* (2015). Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc Natl Acad Sci* **112**: 11371–11376.

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl D a. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–6.

Kuypers MM, Lavik G, Woebken D, Schmid M, Fuchs BM, Amann R, *et al.* (2005). Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc Natl Acad Sci U S A* **102**: 6478–6483.

Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian H, *et al.* (2015). VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3**: 1.

Langille M, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes J, *et al.* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–21.

Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

LaPara TM, Hope Wilkinson K, Strait JM, Hozalski RM, Sadowksy MJ, Hamilton MJ. (2015). The Bacterial Communities of Full-Scale Biologically Active, Granular Activated

Carbon Filters Are Stable and Diverse and Potentially Contain Novel Ammonia-Oxidizing Microorganisms. *Appl Environ Microbiol* **81**: 6864–6872.

Lautenschlager K, Hwang C, Ling F, Liu W-T, Boon N, Köster O, *et al.* (2014). Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. *Water Res* **62**: 40–52.

Lebedeva E V., Off S, Zumbrägel S, Kruse M, Shagzhina A, Lücker S, *et al.* (2011). Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring. *FEMS Microbiol Ecol* **75**: 195–204.

Lees H. (1952). Hydroxylamine as an Intermediate in Nitrification. *Nature* **169**: 156–157.

Leibold MA, McPeek MA. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology* **87**: 1399–1410.

Levy-Booth DJ, Prescott CE, Grayston SJ. (2014). Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biol Biochem* **75**: 11–25.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. (2014). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.

Li H, Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–95.

Lotti T, Kleerebezem R, van Erp Taalman Kip C, Hendrickx TLG, Kruit J, Hoekstra M, *et al.* (2014). Anammox Growth on Pretreated Municipal Wastewater. *Environ Sci Technol* **48**: 7874–7880.

Lücker S, Schwarz J, Gruber-Dorninger C, Spieck E, Wagner M, Daims H. (2014). Nitrotoga-like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants. *ISME J* **9**: 708–720.

Lucker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, *et al.* (2010). A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci* **107**: 13479–13484.

Luo C, Rodriguez-R LM, Konstantinidis KT. (2014). MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res* **42**: 1–12.

Lytle DA, Sorg TJ, Wang L, Muhlen C, Rahrig M, French K. (2007). Biological nitrification in a full-scale and pilot-scale iron removal drinking water treatment plant. *J Water Supply Res Technol - AQUA* **56**: 125–136.

Madsen EL. (2005). Identifying microorganisms responsible for ecologically significant biogeochemical processes. *Nat Rev Microbiol* **3**: 439–446.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

Martiny AC, Albrechtsen H-J, Arvin E, Molin S. (2005). Identification of Bacteria in Biofilm and Bulk Water Samples from a Nonchlorinated Model Drinking Water Distribution System-Nitrospira.pdf. *Appl Environ Microbiol* **71**: 8611–8617.

Meincke M, Bock E, Kastrau D, Kroneck PMH. (1992). Nitrite oxidoreductase from Nitrobacter hamburgensis: Redox centers and their catalytic role. *Arch Microbiol* **158**:

127–131.

Menzel P, Krogh A. (2015). Kaiju : Fast and sensitive taxonomic classification for metagenomics. *bioRxiv* **7**: 1–9.

Mills J, Wyborn NR, Greenwood JA, Williams SG, Jones CW. (1997). An outer-membrane porin inducible by short-chain amides and urea in the methylotrophic bacterium Methylophilus methylotrophus. *Microbiology* **143 (Pt 7**: 2373–2379.

Miyatake T, MacGregor BJ, Boschker HTS. (2009). Linking microbial community function to phylogeny of sulfate-reducing Deltaproteobacteria in marine sediments by combining stable isotope probing with magnetic-bead capture hybridization of 16S rRNA. *Appl Environ Microbiol* **75**: 4927–4935.

Mulder A, Van de Graaf AA, Robertson LA, Kuenen JG. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *Fems Microbiol Ecol* **16**: 177–184.

Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F, *et al.* (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci U S A* **105**: 17861–6.

Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, *et al.* (2010). Use of Ichip for High-Throughput In Situ Cultivation of 'Uncultivable' Microbial Species. *Appl Environ Microbiol* **76**: 2445–2450.

Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, *et al.* (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* **32**: 822–828.

Nitzsche KS, Weigold P, Lösekann-Behrens T, Kappler A, Behrens S. (2015). Microbial community composition of a household sand filter used for arsenic, iron, and manganese removal from groundwater in Vietnam. *Chemosphere* **138**: 47–59.

Nobu MK, Narihiro T, Rinke C, Kamagata Y, Tringe SG, Woyke T, *et al.* (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J* **9**: 1710–1722.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res* gr.213959.116.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **27**: 29–34.

Ouellet H, Ranguelova K, LaBarre M, Wittenberg JB, Wittenberg BA, Magliozzo RS, *et al.* (2007). Reaction of Mycobacterium tuberculosis truncated hemoglobin O with hydrogen peroxide: Evidence for peroxidatic activity and formation of protein-based radicals. *J Biol Chem* **282**: 7491–7503.

Overbeek R, Begley T, Butler RM, Choudhuri J V., Chuang HY, Cohoon M, *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.

Palmer K, Biasi C, Horn MA. (2012). Contrasting denitrifier communities relate to contrasting N2O emission patterns from acidic peat soils in arctic tundra. *ISME J* **6**: 1058–77.

Palomo A, Jane Fowler S, Gülay A, Rasmussen S, Sicheritz-Ponten T, Smets BF. (2016).

Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *ISME J* **10**: 2569–2581.

Peng Y, Leung HCM, Yiu SM, Chin FYL. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–8.

Penton CR, Johnson TA, Quensen JF, Iwai S, Cole JR, Tiedje JM. (2013). Functional genes to assess nitrogen cycling and aromatic hydrocarbon degradation: primers and processing matter. *Front Microbiol* **4**: 1–17.

Petersen TN, Lukjancenko O, Thomsen MCF, Maddalena Sperotto M, Lund O, Møller Aarestrup F, *et al.* (2017). MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads. *PLoS One* **12**: e0176469.

Pinto AJ, Marcus DN, Ijaz Z, Bautista-de los Santos QM, Dick GJ, Raskin L. (2015). Metagenomic Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking Water System. *mSphere* **1**: e00054-15.

Pjevac P, Schauberger C, Poghosyan L, Herbold CW, van Kessel MAHJ, Daebeler A, *et al.* (2016). AmoA-targeted polymerase chain reaction primers for the specific detection and quantification of comammox Nitrospira in the environment. *bioRxiv*. http://biorxiv.org/content/early/2016/12/27/096891.abstract.

Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, *et al.* (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**: D284–D289.

Pratscher J, Dumont MG, Conrad R. (2011). Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. *Proc Natl Acad Sci* **108**: 4170–4175.

Pride DT. (2003). Evolutionary Implications of Microbial Genome Tetranucleotide Frequency Biases. *Genome Res* **13**: 145–158.

Prosser JI, Nicol GW. (2012). Archaeal and bacterial ammonia-oxidisers in soil: The quest for niche specialisation and differentiation. *Trends Microbiol* **20**: 523–531.

Prosser JI, Nicol GW. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931–2941.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, *et al.* (2012). The Pfam protein families database. *Nucleic Acids Res* **40**: D290–D301.

Raghoebarsing A a, Pol A, van de Pas-Schoonen KT, Smolders AJP, Ettwig KF, Rijpstra WIC, *et al.* (2006). A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**: 918–21.

Rho M, Tang H, Ye Y. (2010). FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res* **38**: 1–12.

Rughöft S, Herrmann M, Lazar CS, Cesarz S, Levick SR, Trumbore SE, *et al.* (2016). Community Composition and Abundance of Bacterial, Archaeal and Nitrifying Populations in Savanna Soils on Contrasting Bedrock Material in Kruger National Park, South Africa. *Front Microbiol* **7**: 1–16.

Sandberg R, Bränden C-I, Ernberg I, Cöster J. (2003). Quantifying the species-specificity in genomic signatures, synonymous codon choice, amino acid usage and G+C content. *Gene* **311**: 35–42.

Sawayama S. (2006). Possibility of anoxic ferric ammonium oxidation. *J Biosci Bioeng* **101**: 70–72.

Schott J, Griffin BM, Schink B. (2010). Anaerobic phototrophic nitrite oxidation by Thiocapsa sp. strain KS1 and Rhodopseudomonas sp. strain LQ17. *Microbiology* **156**: 2428–2437.

Schramm A, De Beer D, Van Den Heuvel JC, Ottengraf S, Amann R. (1999). Microscale distribution of populations and activities of Nitrosospira and Nitrospira spp. along a macroscale gradient in a nitrifying bioreactor: Quantification by in situ hybridization and the use of microsensors. *Appl Environ Microbiol* **65**: 3690–3696.

Schramm A, De Beer D, Wagner M, Amann R. (1998). Identification and activities in situ of Nitrosospira and Nitrospira spp. as dominant populations in a nitrifying fluidized bed reactor. *Appl Environ Microbiol* **64**: 3480–3485.

Schubert CJ, Durisch-kaiser E, Wehrli B, Thamdrup B, Lam P, Kuypers MMM. (2006). Anaerobic ammonium oxidation in a tropical freshwater system (Lake Tanganyika). *Env Microbiol* **8**: 1857–1863.

Schwab VF, Hermann M, Roth V-N, Gleixner G, Lehmann R, Pohnert G, *et al.* (2016). Functional diversity of microbial communities in pristine aquifers inferred by PLFA - and sequencing - based approaches. *Biogeosciences Discuss* 1–45.

Sedlar K, Kupkova K, Provaznik I. (2017). Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Comput Struct Biotechnol J* **15**: 48–55.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**: 811–4.

Shen Y-C, Hu Y-N, Shaw G-C. (2016). Expressions of alkaline phosphatase genes during phosphate starvation are under positive influences of multiple cell wall hydrolase genes in Bacillus subtilis. *J Gen Appl Microbiol* **62**: 106–109.

Speth DR, In 't Zandt MH, Guerrero-cruz S, Dutilh BE, Jetten MSM. (2016). Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat Commun* **7**: 11172.

Spieck E, Hartwig C, McCormack I, Maixner F, Wagner M, Lipski A, *et al.* (2006). Selective enrichment and molecular characterization of a previously uncultured Nitrospira-like bacterium from activated sludge. *Environ Microbiol* **8**: 405–415.

Stein LY, Sayavedra-Soto LA, Hommes NG, Arp DJ. (2000).Differential regulation of amoA and amoB gene copies in Nitrosomonas europaea.*FEMS Microbiol Let* **192**:163–168

Strous M, Kraft B, Bisdorf R, Tegetmeyer HE. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front Microbiol* **3**: 1–11.

Sutton M a, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W. (2011). Too much of a good thing. *Nature* **472**: 159–161.

Tatari K. (2014). Nitrification biokinetics in rapid sand filters for drinking water treatment. PhD Thesis, Technical University of Denmark.

Tatari K, Gülay A, Thamdrup B, Albrechtsen H-J, Smets BF. (2017). Challenges in using allylthiourea and chlorate as specific nitrification inhibitors. *Chemosphere* **In press.** e-pub

ahead of print, doi: 10.1016/j.chemosphere.2017.05.005.

Tatari K, Smets BF, Albrechtsen H-J. (2013). A novel bench-scale column assay to investigate site-specific nitrification biokinetics in biological rapid sand filters. *Water Res* **47**: 6380–7.

Tatusov RL. (1997). A Genomic Perspective on Protein Families. *Science 80* **278**:631–637

Tessier AJ, Leibold MA. (1997). Habitat use and ecological specialization within lake Daphnia populations. *Oecologia* **109**: 561–570.

Torge R, Comandini A, Catacchio B, Bonamore A, Botta B, Boffi A. (2009). Peroxidase-like activity of Thermobifida fusca hemoglobin: The oxidation of dibenzylbutanolide. *J Mol Catal B Enzym* **61**: 303–308.

Tringe SG, Hugenholtz P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* **11**: 442–446.

Urfer D, Huck PM, Booth SDJ, Coffey BM. (1997). Biological filtration for BOM and particle removal: a critical review. *J - AWWA* **89**: 83–98.

Vajrala N, Martens-Habbena W, Sayavedra-Soto LA, Schauer A, Bottomley PJ, Stahl DA, *et al.* (2013). Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc Natl Acad Sci U S A* **110**: 1006–11.

Vet WWJM De, Dinkla IJT, Abbas B a, Rietveld LC, Loosdrecht MCM Van. (2012). Gallionella spp. in trickling filtration of subsurface aerated and natural groundwater. *Biotechnol Bioeng* **109**: 904–12.

Villanueva CM, Kogevinas M, Cordier S, Templeton MR, Vermeulen R, Nuckols JR, *et al.* (2014). Assessing exposure and health consequences of chemicals in drinking water: Current state of knowledge and research needs. *Environ Health Perspect* **122**: 213–221.

Vissers EW, Anselmetti FS, Bodelier PLE, Muyzer G, Schleper C, Tourna M, *et al.* (2013). Temporal and spatial coexistence of archaeal and bacterial amoA genes and gene transcripts in lake Lucerne. *Archaea* **2013**. e-pub ahead of print, doi: 10.1155/2013/289478.

Vollmers J, Wiegand S, Kaster A-K. (2017). Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! Rodriguez-Valera F (ed). *PLoS One* **12**: e0169662.

Wagner M. (2009). Single-Cell Ecophysiology of Microbes as Revealed by Raman Microspectroscopy or Secondary Ion Mass Spectrometry Imaging. *Annu Rev Microbiol* **63**: 411–429.

Wang F, Liu Y, Wang J, Zhang Y, Yang H. (2012). Influence of growth manner on nitrifying bacterial communities and nitrification kinetics in three lab-scale bioreactors. *J Ind Microbiol Biotechnol* **39**: 595–604.

Wang J, Xia F, Zeleke J, Zou B, Rhee S. (2016). An improved protocol with a highly degenerate primer targeting genes for community analysis of and. e-pub ahead of print, doi: http://femsec.oxfordjournals.org/.

Wang Y, Ma L, Mao Y, Jiang X, Xia Y, Yu K, *et al.* (2017). Comammox in drinking water systems. *Water Res* **116**: 332–341.

Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U. (1986). Nitrospira marina

gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium.*Arch Microbiol* **144**:1–7

Wessén E, Nyberg K, Jansson JK, Hallin S. (2010). Responses of bacterial and archaeal ammonia oxidizers to soil organic and fertilizer amendments under long-term management. *Appl Soil Ecol* **45**: 193–200.

Wexler M, Bond PL, Richardson DJ, Johnston AWB. (2005). A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* **7**: 1917–26.

White CP, Debry RW, Lytle D a. (2012). Microbial survey of a full-scale, biologically active filter for treatment of drinking water. *Appl Environ Microbiol* **78**: 6390–4.

Williams TJ, Allen MA, DeMaere MZ, Kyrpides NC, Tringe SG, Woyke T, *et al.* (2014). Microbial ecology of an Antarctic hypersaline lake: genomic assessment of ecophysiology among dominant haloarchaea. *ISME J* **8**: 1645–1658.

Wilmes P, Simmons SL, Denef VJ, Banfield JF. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev* **33**: 109–132.

Winogradsky S. (1892). Contributions a la morphologie des organismes de la nitrification. *Arch Sci Biol (St Petersb)* **1**: 86–137.

Wood DE, Salzberg SLS, Venter C, Remington K, Heidelberg J, Halpern A, *et al.* (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**: R46.

Wu ML, Ettwig KF, Jetten MSM, Strous M, Keltjens JT, Niftrik L van. (2011). A new intra-aerobic metabolism in the nitrite-dependent anaerobic methane-oxidizing bacterium *Candidatus* 'Methylomirabilis oxyfera'. *Biochem Soc Trans* **39**: 243–248.

Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, *et al.* (2006). Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A* **103**: 12317–22.

Yamamoto N, Otawa K, Nakai Y. (2010). Diversity and Abundance of Ammonia-Oxidizing Bacteria and Ammonia-Oxidizing Archaea During Cattle Manure Composting. *Microb Ecol* **60**: 807–815.

Yang WH, Weber KA, Silver WL. (2012). Nitrogen loss from soil through anaerobic ammonium oxidation coupled to iron reduction. *Nat Geosci* **5**: 538–541.

Ye L, Shao MF, Zhang T, Tong AHY, Lok S. (2011). Analysis of the bacterial community in a laboratory-scale nitrification reactor and a wastewater treatment plant by 454-pyrosequencing. *Water Res* **45**: 4390–4398.

Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, *et al.* (2002). Nonlinear partial differential equations and applications: Cultivating the uncultured. *Proc Natl Acad Sci* **99**: 15681–15686.

Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L. (2015). High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. *MBio* **6**: e02288-14.

Zhu W, Lomsadze A, Borodovsky M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132–e132.

Ziemann M. (2016). Accuracy, speed and error tolerance of short DNA sequence aligners. *bioRxiv* 53686.

10 Papers

I **Palomo, A.**, Fowler, S.J., Rasmussen, S., Sicheritz-Pontén, T., Smets, B.F. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *ISME J.* **10,** 2569–2581 (2016).

II **Palomo, A.**, Pedersen A.G., Fowler, S.J., Dechesne, A., Sicheritz-Pontén, T., Smets, B.F. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox *Nitrospira*. *Submitted manuscript*.

III Fowler, S.J., **Palomo, A.**, Smets, B.F. Comammox Nitrospira are the dominant ammonia oxidizers in diverse RSF communities. *Submitted manuscript*.

IV **Palomo, A.**, Fowler, S.J., Nemer, I.M., Smets, B.F. Examining differential abundance in rapid sand filter microbial communities after short-term ammonium loading-disturbances. *Manuscript in preparation*.

11 Appendices

I RNA extraction from groundwater-fed rapid sand filters

II Single cell sequencing from a groundwater-fed rapid sand filter

TEXT FOR WWW-VERSION (without papers)

In this online version of the thesis, **paper I-V** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from.

DTU Environment
Technical University of Denmark
Bygningstorvet, Bygning 115
2800 Kgs. Lyngby
Denmark

info@env.dtu.dk.